NICE DSU TECHNICAL SUPPORT DOCUMENT 11: ALTERNATIVES TO EQ-5D FOR GENERATING HEALTH STATE UTILITY VALUES

REPORT BY THE DECISION SUPPORT UNIT

March 2011

John Brazier, Donna Rowen

School of Health and Related Research, University of Sheffield, UK

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent Street Sheffield, S1 4DA

Tel (+44) (0)114 222 0734
E-mail dsuadmin@sheffield.ac.uk

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University.

The DSU is commissioned by The National Institute for Health and Clinical Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme. Please see our website for further information www.nicedsu.org.uk

ABOUT THE TECHNICAL SUPPORT DOCUMENT SERIES

The NICE Guide to the Methods of Technology Appraisal[i] is a regularly updated document that provides an overview of the key principles and methods of health technology assessment and appraisal for use in NICE appraisals. The Methods Guide does not provide detailed advice on how to implement and apply the methods it describes. This DSU series of Technical Support Documents (TSDs) is intended to complement the Methods Guide by providing detailed information on how to implement specific methods.

The TSDs provide a review of the current state of the art in each topic area, and make clear recommendations on the implementation of methods and reporting standards where it is appropriate to do so. They aim to provide assistance to all those involved in submitting or critiquing evidence as part of NICE Technology Appraisals, whether manufacturers, assessment groups or any other stakeholder type.

We recognise that there are areas of uncertainty, controversy and rapid development. It is our intention that such areas are indicated in the TSDs. All TSDs are extensively peer reviewed prior to publication (the names of peer reviewers appear in the acknowledgements for each document). Nevertheless, the responsibility for each TSD lies with the authors and we welcome any constructive feedback on the content or suggestions for further guides.

Please be aware that whilst the DSU is funded by NICE, these documents do not constitute formal NICE guidance or policy.

Dr Allan Wailoo

Director of DSU and TSD series editor.

[i] National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal, 2008 (updated June 2008), London.

Acknowledgements

The authors are grateful to John Cairns, Andrew Lloyd, Elly Stolk and Allan Wailoo for their very helpful comments on an earlier draft.

The DSU thanks the team at NICE, led by Zoe Garrett, for reviewing this document.

The production of this document was funded by the National Institute for Health and Clinical Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

This report should be referenced as follows:

Brazier, J.E., Rowen, D. NICE DSU Technical Support Document 11: Alternatives to EQ-5D for generating health state utility values. 2011. Available from http://www.nicedsu.org.uk

EXECUTIVE SUMMARY

The NICE Guide to the Methods of Technology Appraisal expresses a preference for using the EQ-5D for adult populations to estimate the Quality Adjusted Life Year (QALY) impact of different technologies. Alternative methods to generating health state utility values (HSUVs) will be considered by NICE in place of EQ-5D when EQ-5D data are either unavailable or inappropriate. Unavailability should be established from a systematic search of the literature. Even if EQ-5D data is unavailable, mapping can be used as an alternative to estimate EQ-5D values rather than the alternative methods described in this Technical Support Guide (see TSD 10[1]). Claims that EQ-5D is inappropriate for measuring the Health Related Quality of Life (HRQL) for a patient group or a specific treatment must be supported by evidence.

NICE guidance for alternative methods used to generate HSUVs can be summarised as:

- *Provide supporting argument and evidence* for the choice of alternative methods
- *The descriptions of health states* being valued should be based on validated patient reported measures of HRQL
- *Valuation methods* should be comparable to those used to value the EQ-5D
- *The impact of using alternative methods* on the results of the economic evaluation should be provided and compared to EQ-5D where possible

This Technical Support Document (TSD) examines alternative methods to determine the extent to which they are in accordance with the NICE recommendations. Alternative methods to generate HSUVs include other generic preference-based measures, condition-specific preference-based measures of HRQL, health state vignettes, or the direct valuation of own health (such as using the time trade-off (TTO) elicitation technique). Vignettes not based on standardized and validated measures on HRQL and patient own health state valuations do not meet the NICE Methods Guidance and have a limited role. These methods should only be used where there are no other data based on validated HRQL measures.

Alternative generic preference-based measures and condition-specific preference-based measures of HRQL (CSPBMs) derived from validated measures of HRQL can be used. The

scoring should be based on UK general population values elicited using a choice-based technique, preferably using TTO with the same protocol as the UK EQ-5D valuation.

The development of the measure must be fully described including the way health state descriptions are based on validated measures of HRQL and the valuation methods used and their comparability to those used for EQ-5D. Convincing empirical evidence should detail the properties of the alternative measure such as content validity, construct validity, responsiveness and reliability. Evidence should demonstrate the impact on the QALYs gained from using the alternative method to produce HSUVs rather than EQ-5D.

Future research is recommended to examine the use of 'add-on' dimensions to EQ-5D as an alternative to the use of CSPBMs and to the development of a set of measures acceptable to NICE where EQ-5D is not appropriate. Future research is also recommended to examine the comparability and performance of CSPBMs to EQ-5D.

CONTENTS

TABLES & BOXES

1. INTRODUCTION AND BACKGROUND

1.1. PURPOSE OF THIS TECHNICAL SUPPORT DOCUMENT

The Guide to the Methods of Technology Appraisal (Methods Guide) describes key aspects of analyses submitted to the technology appraisals programme.[2] This Technical Support Document (TSD) is part of a wider initiative to produce a series of TSDs that accompany the Methods Guide. Each TSD describes how to use analytical techniques recommended in the Methods Guide, offer suggestions for analyses for areas not currently covered in the Methods Guide and identify areas that would benefit from further methodological research.

This TSD reviews the alternatives to EQ-5D for generating health state utility values other than mapping (which is dealt with in TSD 10[1]) and provides guidance on how they might be used to generate health state utility values for use in cost effectiveness models in health technology assessments (HTA) submitted to NICE.

1.2. INTRODUCTION

The latest NICE Methods Guidance on methods in Technology Assessment has expressed a preference for using the EQ-5D in adult populations for the estimation of the quality adjusted life year (QALY) impact of different technologies (see TSD 8[3]). The EQ-5D has become a widely used generic preference-based measure of health that has been validated for use in many different conditions.[4] However, the NICE Methods Guide has accepted that there may be circumstances where EQ-5D data are not available or where it is not appropriate in terms of the sensitivity and relevance of the descriptive system to the condition and the consequences of the intervention being assessed. Alternative methods for generating health state utility values are permitted in these situations. The remainder of this section further describes the circumstances where alternative methods can be used. Here we are concerned only with adult populations. For guidance on generating health state utility values for child populations refer to Technical Support Document 1. Section 2 describes and elucidates what the NICE Methods Guide has to say about the use of alternative methods. Section 3 reviews four alternative methods for generating health state utility values (HSUVs) that to varying degrees comply with NICE Methods Guidance: other generic preference-based measures, condition-specific preference-based measures, health state vignettes and the valuation of own health by patients. A set of recommendations are provided at the end.

1.3. WHEN CAN ALTERNATIVE MEASURES BE USED?

1.3.1. When EQ-5D data are not available

The lack of EQ-5D data may arise from a failure to use EQ-5D in key clinical trials. However the lack of data collected in specific trials may not demonstrate a lack of availability of EQ-5D data for populating an economic model. There could be other related studies or routine sources that provide the necessary evidence on the values for the key states used in the model that may be undertaken to support the submission or the values might be identified by a systematic search of the literature. Before concluding that EQ-5D estimates are not available for the health states in an economic model, a systematic search of the literature for EQ-5D is recommended (see TSD 9[5]). Another potential source of EQ-5D estimates comes from applying mapping functions to other validated Health Related Quality of Life (HRQL) measures used in relevant clinical studies or routine sources that can be used to predict EQ-5D values and this is covered in TSD 10[1]. Where there are no means of obtaining EQ-5D estimates for the relevant states, then an alternative method needs to be used.

1.3.2. When EQ-5D is not appropriate

NICE Methods Guide requires evidence to be submitted that demonstrates the EQ-5D is not appropriate for the patient group in terms of psychometric criteria such as validity and responsiveness. It is not sufficient to simply claim EQ-5D is not appropriate. The appropriateness of EQ-5D across conditions has been discussed at some length in TSD 8[3] and so we do not propose to discuss it further here. However, NICE accepts there will be times when EQ-5D is not appropriate and so alternative methods need to be considered.

2. REQUIREMENTS OF NICE METHODS GUIDANCE

NICE Methods Guide allows for alternative methods for generating HSUVs and provides some brief direction for what is likely to be acceptable. This guidance can be summarised under four key headings:

- *Provide supporting argument and evidence* for the choice of alternative methods
- *The descriptions of health states* being valued should be based on validated measures of HRQL
- *Valuation methods* should be comparable to those used to value the EQ-5D
- *The impact of using alternative methods* on the results of the economic evaluation should be provided and compared to EQ-5D where possible

These four areas of guidance are developed further in this TSD and their implications set out for the use of alternative methods for generating HSUVs.

2.1. SUPPORTING EVIDENCE FOR ALTERNATIVE METHOD

The rationale for the choice of the alternative measure must be provided in the submission, along with supporting evidence on its appropriateness for the outcomes of the intervention and the patient group. This is likely to be a natural development of the argument for why the EQ-5D is not appropriate, where the special features of the outcomes for patient quality of life are likely to have been identified. A study comparing the performance of the alternative measure and the EQ-5D in terms of the content validity, construct validity, reliability and responsiveness of the descriptive system would help establish the support for the chosen measure (see TSD 8[3] for full description of terms). For content validity for example, this might involve identifying specific dimensions absent from EQ-5D that are covered by the alternative measure identified from qualitative interviews with patients (such as an energy dimension). To be more convincing, the superior content needs to be empirically shown to result in demonstrable improvements in construct validity, such as reflecting known group differences and/or responsiveness to changes in health. Where a condition-specific preference-based measure is not based on an existing validated patient reported measure but has been developed de novo, the same level of evidence is required.

Any decision by NICE regarding the appropriateness of one measure over another is ultimately a judgement.

2.2. DESCRIPTIONS OF HEALTH STATES

The Methods Guide states that the health states should be based on validated HRQL measures. These should also be patient reported measures in most circumstances, except where they are unable to complete a questionnaire and then they can be completed on their behalf by a carer. The measure must be validated, which would rule out vignettes developed specifically for an economic model using expert opinion as has been common practice in the past.[6] However, it is not clear what being 'based' on a measure of HRQL actually means in practice. There are direct empirical methods discussed in section 3.2 for generating health states from validated measures of HRQL based on the careful selection of representative items, but there are less formal methods that come closer to the vignettes based approach but are nonetheless based on a validated HRQL measure (see section 3.3).

2.3. METHODS OF VALUATION

Methods for valuing health states covers the specific valuation techniques used, how they are used and on whom they are used.

There are a number of valuation techniques. The most widely used techniques for obtaining HSUVs have been Standard Gamble (SG), Time Trade-off (TTO), and Visual Analogue Scale (VAS), and more recently ordinal methods like Best Worst Scaling (BWS). Given that QALYs are the main measure of benefit in cost-effectiveness models for NICE, this would also require that the values generated by the measure are on the full health-dead scale, where full health is one and dead is zero. This would rule out most of the attempts to generate values using ordinal methods since they generate a latent variable that does not lie on the QALY scale, though there is ongoing work to overcome this limitation (for example Salomon (2003)[7] and Ratcliffe *et al* (2009)[8]).

There has been considerable debate in the literature about the most appropriate technique for valuing health states.[9] Different valuation techniques have been shown to generate different values for the same states.[10] The values obtained have also been shown to depend on the specific variant of the technique, such as procedure for arriving at the value of indifference (e.g. by iteration or titration) and mode of administration (e.g. face to face or computer administration). For these reasons comparability with EQ-5D is enhanced by using the same

valuation technique, the same variant of the technique and by the same mode. This is why the NICE Methods Guidance prefers one technique to be used, namely TTO anchored at full health. Full consistency would currently require interview administration using the York MVH TTO protocol.[11] This may change in the future, so might the way states valued worse than dead are handled.[12] Where TTO data are not available, then the Guidance indicates another choice-based method like SG is preferred over VAS.

The source of values has important implications for the value obtained. Patients, for example, tend to generate higher values than members of the general public for the same health states (at least for physical health).[13,14] The main argument for using the general public to value health states hinges on the view that in a publicly funded health care system it is society's resources that are being allocated, and therefore it is the views of the general population that are relevant.[15] While there are arguments for using patient values, the requirement for general population values is quite explicit in the NICE Methods Guidance.

It is important to obtain a value set that is representative of the general public of the UK. Background characteristics have been found to have an impact on HSUVs, such as age and own health, although less than the health state descriptions themselves.[16,17] There are a range of methods for obtaining a representative sample of the UK general population. Ideally there would be a systematic sampling of addresses around the country using a rigorous sampling frame rather than a simple convenience sample. The latter may result in a less than representative sample and so adjustments may need to be made to the weighting of the samples or statistical adjustments using covariates reflecting the influence of different background characteristics. Any adjustments will be limited to measured characteristics in a valuation survey, however there may be other factors not taken into account (such as arises from using internet samples).

2.4. IMPACT OF ALTERNATIVE METHODS

The NICE Methods Guidance asks for evidence indicating the likely impact on the results of using the alternative measure to calculate QALYs. The Guidance seems to see the alternative measure as being additional to EQ-5D evidence; however this may not be possible where EQ-5D data are not available. Even in the absence of EQ-5D data, it may nonetheless be possible to indicate the likely impact, such as the known differences in values generated by some of

the generic measures and EQ-5D or the fact that patient values tend to be higher than those of the general public. These implications are discussed in the next section.

3. REVIEW OF ALTERNATIVE METHODS

This section reviews the main alternative methods for generating HSUVs that have been submitted to NICE. A review undertaken between 2004 and 2008 of submissions and assessments prior to the latest NICE Methods Guidance found that non-EQ-5D HSUVs SUV accounted for nearly half of the studies conducted for NICE. The HSUVs came from either other generic preference-based measures, health state vignettes, or the direct valuation of own health (such as by TTO or SG).[6] These alternatives will be reviewed, along with the emerging approach of generating condition-specific preference-based measures.

3.1. OTHER GENERIC MEASURES OF HEALTH

The EQ-5D is one of a number of generic preference-based measures of health developed for use in economic evaluation. Generic preference-based measures have two components: firstly a system for describing health or its impact on quality of life using a standardised descriptive system; secondly an algorithm for assigning values to each state described by the system. A health state descriptive system is composed of a number of multi-level dimensions that together describe a universe of health states (such as the EQ-5D that has 5 dimensions with three levels each, generating 3^5= 243 health states). Generic instruments have been developed for use across all patient groups by focusing on core aspects of health. The self-complete questionnaires used to collect health state data are easily included in clinical studies or routine data collection systems with little respondent burden. The valuation component comes from existing preference weights that have been obtained using one of the health state valuation techniques using samples of the general population.

The most widely used generic preference-based measures, other than EQ-5D, are HUI2 and HUI3,[18,19] 15D[20], AQoL[21] and AQoL 2,[22] and SF-6D.[23,24] Whilst these measures all claim to be generic, they differ considerably in terms of the content and size of their descriptive system, the methods of valuation and the populations used to values the health states (differing in country and methods of sampling of the general population). A summary of the

main characteristics of these generic preference-based generic measures of health is presented in Table 1.

3.1.1. Comparison of measures

The EQ-5D is the smallest of the measures, with the others having more dimensions and levels (though the new 5-level version of EQ-5D will to some extent change this). The SF-6D covers similar concepts of HRQL to EQ-5D in terms of physical, mental and social health, but has more detail on some aspects of these dimensions and includes the additional dimension of energy. HUI2 and HUI3 are quite different in content in that they focus on impairments rather than HRQL and so cover hearing, vision, dexterity and so forth. The developers of HUI argue that the HRQL impact is captured in the valuation task. The AQoL and 15D cover more dimensions, with the former achieving this by using a hierarchical structure.

Table 1 summarises the valuation methods used in terms of the valuation technique. The EQ-5D and AQOL were valued using time trade-off (TTO) (though using different variants of the techniques); HUI2, HUI3 and SF-6D were valued using the same variant of standard gamble (SG) but the Canadian valuations of the HUI2 and HUI3 were estimated using mapping functions converting visual analogue (VAS) values into SG, and 15D was valued using the VAS. EQ-5D, SF-6D and HUI-2 have UK value sets, while the HUI-3, 15D and AQOL do not. EQ-5D and SF-6D have value sets from several countries and these have been shown to be significantly different, hence the requirement of NICE that submissions use values obtained from UK general populations.

These differences in descriptive content and methods of valuation have resulted in only poor to moderate agreement between measures (around 0.3-0.5 as measured by the intra class correlation coefficient).[4] Whilst differences in mean scores have often been found to be little more than 0.05 between SF-6D, EQ-5D and HUI3, this mean statistic masks considerable differences in the distribution of scores. Overall the SF-6D, for example, produces relatively lower values for patients with more mild problems and relatively higher values in those with more serious health problems. It would seem the EQ-5D suffers from ceiling effects, with large numbers of respondents reporting no problems (i.e. state 11111), but conversely the SF-

6D suffers from floors effects with a lower limit of 0.301 compared to -0.594 for EQ-5D (see for example Brazier *et al*[25] and Longworth *et al*[26]) Comparisons between HUI3 and EQ-5D have typically found less consistency, with HUI3 sometime being higher and at other times lower (for example Bosch and Hunink[27] and Barton et al[27,28]).

3.1.2. Implications

Given these differences NICE has a preference for one generic measure with the purpose of achieving consistency between studies and health care programmes. However, where EQ-5D is not available or not appropriate, these other generic measures may have a role in providing values for cost effectiveness models for NICE. They are standardised patient reported measures that continue to be used in clinical studies and may be the best available in some circumstances. Furthermore, they are less prone to the problems identified with condition-specific measures, such as the focusing effects described later in this section. Where the descriptive system of another generic can be shown to be more relevant to the condition than EQ-5D, the alternative generic measure may be accepted by NICE. This was the case for Macular Degeneration, where NICE used values generated by the HUI3 to appraise cost-effectiveness (TA155[6]). The problem for NICE is that the valuation methods are not the same and some allowance needs to be made for this. The SF-6D, for example, has been shown to estimate smaller differences than EQ-5D due to its more limited range.[29] However, there is no basis for making a formal adjustment in many cases.

Table 1: Summary of generic preference-based measures of health

Instrument	*Dimensions*	*Severity levels*	*Health states*	*UK values*	*Valuation technique*
AQoL	Independent living (self-care, household tasks, mobility) social relationships (intimacy, friendships, family role), physical senses (seeing, hearing, communication), psychological well-being (sleep, anxiety and depression, pain)	4	16.8 million	No	TTO
AQoL 2	Social (including work, family and intimate relationships) independent Living, mental Health, illness (including pain), values and beliefs, sensory items. Obtained from 20 items	4-6 per item	64 billion	No	TTO
EQ-5D	Mobility, self-care, usual activities, pain/discomfort, anxiety/depression	3	243	Yes	TTO
HUI2	Sensory, mobility, emotion, cognition, self-care, pain (original version contained a fertility dimension for use in a specific patient group)	4-5	8,000	Yes	VAS transformed into SG
HUI3	Vision, hearing, speech, ambulation, dexterity, emotion, cognition, pain	5-6	972,000	No	VAS transformed into SG
QWB	Mobility, physical activity, social functioning 27 symptoms/problems	3 2	945	No	VAS
SF-6D	Physical functioning, role limitation, social functioning, pain, energy, mental health	4-6	18,000 (SF-36) 7,500 (SF-12)	Yes	SG
15D	Mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort/symptoms, depression, distress, vitality, sexual activity	4-5	31 billion	No	VAS

3.2. CONDITION-SPECIFIC MEASURES

As described above for generic preference-based measures, condition-specific preference-based measures (CSPBM) also consist of two components: a descriptive system for describing health or its impact on quality of life; and an algorithm for assigning health state values to every state defined by the descriptive system. Table 2 presents a summary of some condition-specific preference-based measures for use in adult populations. These measures have not been approved by NICE, but are presented for illustration of this class of measure. Many condition-specific measures describe symptoms or symptom bother, rather than HRQL. Preference-based measures derived from measures such as the Index of Erectile Function or the Menopause-specific quality of life questionnaire, for example, would not meet the NICE reference case for this reason. Others have values obtained using a non-preference based technique such as VAS. There are also condition-specific preference-based measures that are not anchored onto the 1-0 full health-dead scale and since these cannot be used to generate QALYs they are not included in the list. Whether or not CSPBMs would be acceptable to NICE depends on the content of the descriptive system and the method of valuation, and these are now considered in more detail.

3.2.1. Development of health states

Unlike generic preference-based measures, the descriptive system for a CSPBM is typically derived from an established patient-report condition-specific measure that does not contain preferences in its scoring. The descriptive system contains a subset of items from the original measure that are representative of the original measure. The items are typically chosen using psychometric analysis of existing data using standard psychometric criteria and techniques such as factor and Rasch analysis. A rigorous approach outlined in the literature[30] and applied to derive the classification systems of several CSPBMs[30-32] involves 5 steps: (1) establish dimensions; (2) eliminate items per dimension; (3) select items per dimension; (4) explore reducing severity levels per dimension; (5) validate steps (1) to (4) on other datasets. Step 1 is conducted using factor analysis and steps 2-4 are conducted using Rasch analysis and standard psychometric analysis. This approach ensures that the classification system best represents the dimensionality of the original measure and that the selected items that perform

the best psychometrically are chosen. It is recommended that this process is used to derive classification systems from existing validating measures.

To apply these measures, the original self-complete questionnaire must be used to collect the health state data. For example, utilities for overactive bladder is obtained by: firstly including the self-complete Overactive Bladder Questionnaire (OABq) in a trial; and secondly applying an algorithm to convert OABq responses into health states on the preference-based measure and assign the appropriate HSUV.

Two of the condition-specific measures are entirely new measures (for vision/visual impairment and Parkinson's disease), meaning that a comprehensive description of their construction must be provided if they are used to generate utility values. For the development of the classification system refer to the helpful guidelines outlined in the US FDA Guidance on the development of patient-reported outcome measures to support labelling claims.[33] Note that the scoring of the measure should be in accordance with the guidelines below. New measures should be validated and tested with an emphasis on content validity, construct validity, responsiveness and reliability.

The rigorous application of psychometric and qualitative methods to generate the descriptive systems is necessary but not sufficient. It is also important that the content focuses on the impact of the condition and its treatment on HRQL rather than purely the symptoms.

3.2.2. Valuation

CSPBMs use a variety of techniques to elicit values for the descriptive system including SG, TTO, VAS and combinations of these. VAS is not a choice-based method and its disadvantages are well cited in the literature (see Brazier *et al*[4] for an overview). For this reason utility scores based on VAS values alone (the measure for lung cancer) are regarded as weaker evidence.

It has been found in the literature that utilities elicited for the same health state vary both by elicitation technique (for example Tsuchiya *et al*,[10] by valuation protocol (for example Dolan[34]) and by country (for example Yang *et al*[35]). This raises issues of comparability of the

preferred EQ-5D utilities to utilities generated via alternative measures. Therefore NICE guidance states a preference for measures that use methods of valuation that are comparable to the method used to produce the UK EQ-5D value set. This method involves the use of the MVH protocol of the TTO using a visual prop and face-to-face interviews to elicit utility values from members of the UK population.[11] A large proportion of the measures meet this requirement (asthma, cancer, menopause, overactive bladder, pulmonary hypertension). Two additional measures meet part of the requirement: one measure uses TTO to elicit values using an alternative protocol on members of the Dutch general population (erectile (dys)function); another measure uses SG to elicit utilities from the UK general population (urinary incontinence).

NICE stipulates that utilities should be based on public preferences, meaning that the two measures with value sets derived using patient values (vision/visual impairment and Parkinson's disease) are not in accordance with the NICE Methods Guide.

Most preference-based measures define a large number of health states and it is unfeasible to elicit values for every state from the general population. General population values are elicited for a sample of health states and HSUVs are estimated for all other health states defined by the classification system using modeling. The modeling techniques are the same as those used for the generic preference-based measures (see Brazier *et al* [4] for an overview). The type of modeling varies depending on the valuation technique, classification system and data. The model specification can take a number of forms depending on which best suits the data.

3.2.3. Implications

CSPBMs have also been criticized as being susceptible to focusing effects in the valuation stage of the development of the measure. General population HSUVs are elicited by asking respondents how good or bad they think living in a health state would be. The respondent will focus on the health state description provided and may exaggerate the importance of the problems mentioned in the description, meaning that the utility decrement associated with the condition may be exaggerated. This is likely to be more of a problem with descriptions of symptoms rather than the impact on HRQL.

CSPBMs have also been criticized for their inability to capture side effects and co-morbidities. Some CSPBMs have a narrow focus, which can be seen by examining the dimensions included in the descriptive systems of CSPBMs (Table 2). For example the menopause-specific measure includes only dimensions capturing menopause symptoms and related quality of life and hence will be unable to capture the impact of any co-morbidities or side-effects of treatment. In contrast some measures have a much wider focus, such as the measures for cancer and lung cancer, and may be able to capture co-morbidities and side-effects. The ability to capture these wider effects is therefore likely to vary by measure, and it is recommended that supporting evidence is provided if it is claimed that the measure is able to capture co-morbidities and side-effects.

When CSPBMs are used to generate QALY estimates analyses using empirical data to examine the performance of the CSPBM should be provided. The analyses should examine the properties of the measure including its content validity, construct validity, responsiveness and reliability. Details should also be provided of the development of the descriptive system, valuation and methods used to produce the value set. Evidence should be provided demonstrating how the use of the CSPBM rather than EQ-5D to produce utilities has impacted on the estimation of QALYs gained. However it is recognized that empirical evidence may not always be feasible, for example if the EQ-5D is inappropriate for a population there will be few or no studies in the patient population of interest using the EQ-5D. Under these circumstances the evidence may instead focus on the inappropriateness of EQ-5D for that population and the properties of the CSPBM that is used.

CSPBMs meet the NICE Methods Guidance for alternatives to EQ-5D provided they are based on validated measures of HRQL using valuation methods comparable to those used for the UK EQ-5D value set.

Table 2: Selection of condition-specific preference-based measures of health

Condition	*Measure derived from*	*Dimensions*	*Severity levels*	*Health states*	*UK values*	*Valuation technique*
Asthma[32,36]	Asthma Quality of Life Questionnaire (AQLQ)	Concern, short of breath, weather and pollution, sleep, activities	5	3125	Yes	TTO
Cancer[31]	EORTC QLQ-C30	Physical functioning, role functioning, pain, emotional functioning, social functioning, fatigue and sleep disturbance, nausea, constipation and diarrhoea	4-5	81920	Yes	TTO
Erectile (dys)functioning[37]	Index of Erectile Function (IIEF)	Ability to attain an erection sufficient for satisfactory sexual performance, ability to maintain an erection sufficient for satisfactory sexual performance	5	25	No	TTO
Lung cancer[38,39]	FACT-L	Physical, social/family, emotional, functional, symptoms - general: symptoms – specific	2	64	Yes	VAS
Menopause [40]	Menopause-specific quality of life questionnaire	Hot flushes, aching joints/muscles, anxious/frightened feelings, breast tenderness, bleeding, vaginal dryness, undesirable androgenic signs	3-5	6075	Yes	TTO
Overactive bladder[30,41]	Overactive bladder questionnaire (OABq)	Urge to urinate, urine loss, sleep, coping, concern	5	3125	Yes	TTO
Parkinson's disease[42]	N/A	Disease severity, proportion of the day with ‘off-time’ (impact on quality of life due to condition covering domains: social function, ability to carry out daily activities, psychological function)	2-5	10	No	VAS and SG

Condition	*Measure derived from*	*Dimensions*	*Severity levels*	*Health states*	*UK values*	*Valuation technique*
Pulmonary hypertension[43]	Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR)	Social activities, travelling, dependence, communication	2-3	36	Yes	TTO
Urinary incontinence[44]	The King's Health Questionnaire	Role limitations, physical limitations, social limitations/family life, emotions, sleep/energy	4	1024	Yes	SG
Vision/visual impairment[45-47]	N/A	Physical well-being, independence, social well-being, emotional well-being, self-actualization, planning and organization	5-7	45,360	No	VAS and TTO

3.3. VIGNETTES

This approach involves constructing a vignette or scenario to describe each of the frequently occurring states associated with a condition and its treatment for respondents to value in the usual way. The vignettes are usually based on interviews with patients and professionals. They can incorporate a range of information about the impact of the condition and its treatment. Vignettes can take the form of a text narrative (for example Sackett and Torrance[48]) or more structured descriptions using a bullet point format in a similar way to the generic measures (for example Bass *et al*[49]). Researchers have explored alternative narrative formats, such as the use of videos or simulators.[50,51]

The valuation of a few vignettes provides a convenient way to obtain mean values for the set of severity states around which a cost effectiveness model is structured and was used quite commonly in NICE submissions prior to the latest Methods Guide.[6] They have the advantage of being comparatively easy and quick to prepare, and can be prepared with little or no patient level data though such data would be expected to improve them. There may be some key states where it is not possible for practical or ethical reasons to collect HRQL data. The states can be designed to incorporate concerns of importance to patients with the condition and to include the side-effects of treatment. However as they are typically specific to a condition and its treatment and tend to be valued using different methods, they further reduce the comparability of technology appraisal submissions to NICE across different treatments and patient groups.

The validity of the vignettes depends on the rigour with which they are designed. Where extensive qualitative work is undertaken with patients to construct the vignettes using techniques such as in-depth interviews and focus groups then this provides some evidence for the validity of the descriptions. Their credibility will also be enhanced where the descriptions have been verified by an independent group. Methods have been developed for constructing vignettes using pertinent HRQL data from clinical studies and this provides some quantitative bases for their construction.[52]

The problem is that no matter how good the qualitative work, the vignettes will not be able to fully reflect outcomes experienced by patients in a given vignette state. Within a given clinical state such as one year post hip fracture, for example, patients experience a varied distribution of symptoms, physical functioning, pain and feelings of well-being that will not be properly captured. The extent to which a set of vignettes accurately represent these distributions of states is a quantitative matter that is not usually examined, but will be limited by the small number of vignettes usually constructed. By contrast, the distribution of outcomes can be captured by standardised descriptive systems since they are administered to patients experiencing the states and mean values and variances estimated.

Vignettes that have not been based on HRQL data do not meet the NICE Methods Guidance for alternatives to EQ-5D. However, vignettes may have a limited role where there are no data available using validated HRQL measures.

3.4. PATIENTS' OWN HEALTH STATE VALUATION

This is where the patient is asked to value his or her own health using a preference elicitation technique such as TTO or SG. This differs fundamentally from the other approaches reviewed so far where respondents (whether or not they have the condition) are asked to value hypothetical states selected by the researcher. This approach avoids the need to describe a state of health, since the patient is experiencing it. This has the attraction of avoiding all the problems of poor coverage, insensitivity and lack of meaning associated with many health state descriptive systems.

There are technical and ethical obstacles to collecting health state valuation data from patients that ask them life and death questions, such as TTO and SG. However, the main problem in the context of a NICE submission is that they are not the same as general population values. The valuation of health states by patients in those states have been found to be consistently higher than those obtained from respondents asked to imagine health states, with some evidence suggesting their values are lower than those of the general population for mental health states.[53] Again, where the EQ-5D has been shown to be inappropriate and there are no standardized validated HRQL questionnaires, then there may be a role for these types

of data. The impact of this would again need to be weighed carefully, but the size and even the direction of any difference will be specific to the states.

4. CONCLUSIONS

NICE expressed a preference for using the EQ-5D in adult populations to estimate the QALY impact of different technologies in the Methods Guide for Technology Appraisal. This is to enable comparability of interventions across different technologies and patient groups. Whilst it is recognized that EQ-5D is not always available in key studies or appropriate for all patient populations, the use of alternative measures or methods to estimate QALYs reduces comparability across interventions and patient groups. For this reason NICE alternative measures or methods should only be used in place of EQ-5D to estimate QALYs when: 1) EQ-5D is appropriate but a systematic search of the literature found that it was *unavailable*; or 2) EQ-5D has been found to be *inappropriate* for that patient population and this can be demonstrated using *evidence*. The usage of alternative measures can be reduced by ensuring that EQ-5D data is collected in key studies and trials. Yet as NICE is only one of the many organizations and bodies that require specific evidence to be collected this will remain a barrier to collecting EQ-5D data. However, mapping is one alternative that can be used to estimate EQ-5D utilities even when EQ-5D was not included in the trial or study. For comparability purposes the use of mapped EQ-5D values is preferred to the use of alternative measures where EQ-5D is appropriate but unavailable (see TSD 9[5] for further information).

There are several alternative methods that can be used to estimate QALYs of differing acceptability under the NICE Methods Guide. The alternatives most likely to be accepted by NICE are preference-based measures derived from validated measures of HRQL, with the value set obtained from the general population preferably using techniques similar to the protocol used to obtain the UK EQ-5D value set. Any new measures should be validated. Empirical evidence regarding the preference-based measure should be provided to enable the decision-maker to determine the impact of using the measure in terms of the comparability, credibility, reliability and validity of the QALY estimates.

4.1. RECOMMENDATIONS FOR FURTHER RESEARCH

More specific guidance is needed outlining the conditions and patient groups where EQ-5D is appropriate and inappropriate to inform researchers of which measures should be included in key trials (see TSD 8[3]). In order to meet these requirements further research is needed on the appropriateness of EQ-5D across conditions and patient groups in the form of reviews of existing evidence and where necessary further primary studies into the content validity, construct validity and responsiveness of EQ-5D. This research needs to incorporate the new 5-level version of the EQ-5D. Where EQ-5D is not appropriate research is required into appropriate alternative generic or CSPBMs of HRQL for such conditions. Comparability within a condition at least would be enhanced by having guidance on the best alternative. Given the problems of CSPBMs (e.g. focusing effects), then another solution is worth pursuing which is the use of 'add-on' dimensions to the EQ-5D. For example, if evidence demonstrates that the EQ-5D does not capture a dimension of key importance to a patient group, the dimension can be 'added-on' to the EQ-5D classification system and appropriate value set obtained. Another potential solution in the long term would be a new generic measure based on a definition of quality of life or well-being that was agreed by NICE that was relevant to all conditions.

Whilst there is a large literature examining the performance and comparability of the generic preference-based measures, there is relatively little research on CSPBMs. Future research is recommended to examine the comparability and performance of CSPBMs in relation to EQ-5D. This will inform researchers and policy makers of the implications of using CSPBMs.

5. SUMMARY OF RECOMMENDATIONS

When alternative measures should be used in place of EQ-5D

- Alternative methods can only be used in place of EQ-5D when it is unavailable in a relevant patient group or the EQ-5D is inappropriate for the patient group.
- A systematic search of the literature should be undertaken to determine whether EQ-5D data is unavailable.

- Evidence should be provided to support claims that EQ-5D is inappropriate in terms of psychometric criteria such as validity and responsiveness.

Types of alternative methods that can be used

- Vignettes and patient own health state valuations do not meet the NICE Methods Guidance for alternatives to EQ-5D. These only have a role where there are no data from validated HRQL measures.
- Alternative generic preference-based measures can be used as an alternative to EQ-5D using UK general population value sets.
- CSPBMs can be used as an alternative to EQ-5D provided they are based on validated measures of HRQL and use valuation methods comparable to those used for the UK EQ-5D value set on a representative sample of the UK general population. An outline of the development of the measure should be provided.

Consideration of impact and evidence when using alternative methods

- Quantification of the impact on QALYS gained from using the alternative measure (or method) to produce utilities rather than EQ-5D

Box 1

1. *Supporting argument*: reasons for choice of alternative and supporting evidence for chosen measure must be submitted
2. *Descriptions of states*: health state descriptions should be based on validated HRQL measures.
3. *Valuation methods*: comparable to those used to value the EQ-5D (i.e. TTO with full health as upper anchor and general population values).
4. *Impact of using alternative*: indicate any evidence that will help the committee understand the extent to which the choice of measure impacts on the results.

Box 1: Requirements of alternative methods by NICE Methods Guidance

6. REFERENCES

1. Longworth, L., Rowen, D. NICE DSU Technical Support Document 10: The use of mapping methods to estimate health state utility values. 2011; available from http://www.nicedsu.org.uk/index.htm

2. National Institute of Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. 2008; London, NICE.

3. Brazier, J.E., Longworth, L. NICE DSU Technical Support Document 8: An introduction to the measurement and valuation of health for NICE submissions. 2011; available from http://www.nicedsu.org.uk/index.htm

4. Brazier, J.E., Ratcliffe, J., Solomon, J.A., Tsuchiya, A. Measuring and valuing health for economic evaluation. Oxford University Press, Oxford; 2007.

5. Papaioannou, D., Brazier, J.E., Paisley, S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. 2011; available from http://www.nicedsu.org.uk/index.htm

6. Tosh, J., Longworth, L., George, E. Utility Values in NICE Technology Appraisals. *Value in Health* 2011; 14 (1): 102-109.

7. 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* 2003; 1(1):12.

8. Ratcliffe, J., Brazier, J., Tsuchiya, A., Symonds, T., Brown, M. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. *Health Economics* 2009; 18(11):1261-1276.

9. Green, C., Brazier, J., Deverill, M. Valuing health-related quality of life. A review of health state valuation techniques. *PharmacoEconomics* 2000; 17:151-165.

10. Tsuchiya, A., Brazier, J., Roberts, J. Comparison of valuation methods used to generate the EQ5D and the SF6D value sets in the UK. *Journal of Health Economics* 2006; 25:334-346.

11. Dolan, P. Modeling valuations for EuroQol health states. *Medical Care* 1997; 35(11):1095-1108.

12. Devlin, N., Tsuchiya, A., Buckingham, K.J., Tilling, C. A Uniform Time Trade Off Method for States Better and Worse than Dead: Feasibility Study of the 'Lead Time' Approach. *Health Economics* 2011; 20 (3): 348-361.

13. Boyd, N.F., Sutherland, H.J., Heasman, Z.J., Cummings, B.J. Whose values for decision making? *Medical Decision Making* 1990; 10:58-67.

14. Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C. et al. Should patients have a greater role in valuing health states? *Applied Health Economics and Health Policy* 2005; 4:201-208.

15. Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C. Cost-effectiveness in health and medicine. Oxford University Press, Oxford; 1996.

16. Dolan, P., Roberts, J. To what extent can we explain time trade-off values from other information about respondents? *Social Science and Medicine* 2002; 54(6):919-929.

17. Kharroubi, S.A., Brazier, J.E., Roberts, J., O'Hagan, A. Modelling SF-6D health state preference data using a nonparametric Bayesian method. *Journal of Health Economics* 2007; 26(3):597-612.

18. Feeny, D., Furlong, W., Torrance, G.W., Goldsmith, C.H., Zhu, Z., Depauw, S. et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care* 2002; 40(2):113-128.

19. Torrance, G.W., Feeny, D.H., Furlong, W.J., Barr, R.D., Zhang, Y., Wang, Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care* 1996; 34(7):702-722.

20. Sintonen, H., Pekurinen, M. A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In: Anonymous, eds. *Quality of life assessment: key issues in the 1990s*. Kluwer Academic Publishers; Dordecht, Netherlands: 1993; 185-195.

21. Hawthorne, G., Richardson, J., Osbourne, R. The Assessment of Quality of Life (AQoL) instrument: A psychometric measure of health-related quality of life. *Quality of Life Research* 1999; 8(3):209-224.

22. Richardson, J., Peacock, S., Iezzi, A., Day, N.A., Hawthorne, G. Construction and Validation of the Assessment of Quality of Life (AQoL) Mark II Instrument. Centre for Health Economics Research Paper 24, Monash University 2007.

23. Brazier, J., Roberts, J., Deverill, M. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* 2002; 21(2):271-292.

24. Brazier, J.E., Roberts, J. The estimation of a preference-based measure of health from the SF-12. *Medical Care* 2004; 42(9):851-859.

25. Brazier, J., Roberts, J., Tsuchiya, A., Busschbach, J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics* 2004; 13(9):873-884.

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

27. Bosch, J.L.H., Hunink, M. Comparison of the Health Utilities Intdex mark 3 (HUI3) and the EuroQol EQ-5D in patients treated for intermittent claudication. *Quality of Life Research* 2000; 9:591-601.

28. 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 & Health Policy* 2004; 3(2):103-105.

29. Feeny, D., Wu, L., Eng, K. Comparing short form 6D, standard gamble, and health utilities index Mark 2 and Mark 3 utility scores: Results from total hip arthroplasty patients. *Quality of Life Research* 2004; 13:1659-1670.

30. Young, T., Yang, Y., Brazier, J.E., Tsuchiya, A., Coyne, K. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Quality of Life Research* 2009; 18(2):253-265.

31. Rowen, D., Brazier, J., Young, T., Gaugris, S., Craig, B.M., King, M.T. et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value in Health* 2011; Forthcoming.

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

33. U.S.Department of Health and Human Services Food and Drug Administration (FDA). Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. Maryland, FDA.

34. Dolan, P. Aggregating health state valuations. *Journal of Health Services & Research Policy* 1997; 2(3):160-165.

35. Ferreira, L.N., Ferreira, P.L., Rowen, D., Brazier, J. Do Portuguese and UK health state values differ across valuation methods? *Quality of Life Research* 2011; 20 (4): 609-619.

36. Yang, Y., Brazier, J., Tsuchiya, A., Young, T. Estimating a Preference-Based Index for a 5-Dimensional Health State Classification for Asthma Derived From the Asthma Quality of Life Questionnaire. *Medical Decision Making* 2011; 31(2):281-291.

37. Stolk, E.A., Busschbach, J. Validity and feasibility of the use of condition-specific outcome measures in economic evaluation. *Quality of Life Research* 2003; 12(4):363-371.

38. Kind, P., Macran, S. Eliciting social preference weights for functional assessment of cancer therapy-lung health states. *PharmacoEconomics* 2005; 23(11):1143-1153.

39. Lamers, L.M., Uyl-de Groot, C.A., Buijt, I. The use of disease-specific outcome measures in cost-utility analysis: The development of Dutch societal preference weights for the FACT-L scale. *PharmacoEconomics* 2007; 25(7):591-603.

40. Brazier, J.E., Roberts, J., Platts, M., Zoellner, Y.F., Brazier, J.E., Roberts, J. et al. Estimating a preference-based index for a menopause specific health quality of life questionnaire. *Health & Quality of Life Outcomes* 2005; 3:13.

41. Yang, Y., Brazier, J., Tsuchiya, A., Coyne, K. Estimating a preference-based single index from the overactive bladder questionnaire. *Value in Health* 2009; 12(1):159-166.

42. Palmer, C.S., Schmier, J., Snyder, E., Scott, B. Patient preferences and utilities for 'off-time' outcomes in the treatment of Parkinson's disease. *Quality of Life Research* 2000; 9(7):819-827.

43. McKenna, S.P., Ratcliffe, J., Meads, D.M., Brazier, J.E. Development and validation of a preference based measure derived from the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) for use in cost utility analyses. *Health & Quality of Life Outcomes* 2008; 6:65.

44. Brazier, J.E., 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* 2008; 28(1):113-126.

45. Misajon, R., Hawthorne, G., Richardson, J., Barton, J., Peacock, S., Iezzi, A. et al. Vision and quality of life: the development of a utility measure. *Investigative Ophthalmology & Visual Science* 2005; 46(11):4007-4015.

46. Peacock S, Misajon, R., Iezzi, A., Richardson, J., Hawthorne, G., Keeffe, J. Vision and quality of life: Development of methods for the VisQoL vision-related utility instrument. *Ophthalmic Epidemiology* 2008; 15(4):218-223.

47. Richardson, J., Iezzi, A., Sinha, K., Khan, M., Peacock, S., Hawthorne, G. et al. AQoL-7D (Vision) Instrument: Overview, survey results and utility algorithms. *Centre for Health Economics Research Paper 45, Monash University, Australia* 2009.

48. Sackett, D.L., Torrance, G.W. The utility of different health states as perceived by the general public. *Journal of Chronic Diseases* 1978; 31(11):697-704.

49. Bass, E.B., Steinberg, E.P., Pitt, H.A., Griffiths, R.I., Lillemoe, K.D., aba, G.P. et al. Comparison of the rating scale and the standard gamble in measuring patient preferences for outcomes of gallstone disease. *Medical Decision Making* 1994; 14:307-314.

50. Lenert, L.A., Sturley, A.P., Rapaport, M.H., Chavez, S., Mohr, P.E., Rupnow, M. et al. Public preferences for health states with schizophrenia and a mapping function to estimate utilities from positive and negative symptom scale scores.[erratum appears in Schizophr Res. 2005 Dec 1;80(1):135-6]. *Schizophrenia Research* 2004; 71(1):155-165.

51. Aballea, S., Tsuchiya, A. Seeing and doing: Feasibility study towards valuing visual impairment using simulation spectacles. *Health Economics* 2004; 16:537-543.

52. Stein, K., Dyer, M., Crabb, T., Milne, R., Round, A., Ratcliffe, J. et al. An Internet "Value of Health" Panel: recruitment, participation and compliance. *Health and Quality of Life Outcomes* 2006; 4:90.

53. Brazier, J. Measuring and valuing mental health for use in economic evaluation. *Journal of Health Services Research and Policy* 2008; 13(SUPPL. 3):70-75.