

**NICE DSU TECHNICAL SUPPORT DOCUMENT 8:
AN INTRODUCTION TO THE MEASUREMENT AND
VALUATION OF HEALTH FOR NICE SUBMISSIONS**

REPORT BY THE DECISION SUPPORT UNIT

August 2011

John Brazier¹ and Louise Longworth²

1. School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, Sheffield, UK. S1 4DA1.
2. Health Economics Research Group (HERG), Brunel University, Uxbridge, Middlesex, UK. UB8 3PH.

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 Appraisalⁱ 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 DSU thanks Andrew Lloyd, Nancy Devlin, Allan Wailoo and the team at NICE, led by Janet Robertson, for reviewing this document. The editor for the TSD series is Allan Wailoo.

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., 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>

EXECUTIVE SUMMARY

NICE requires consistency in its decision making and so has developed a reference case of methods for Technology Appraisal that includes the measurement and valuation of health. This TSD starts with an overview of the reference case, and then goes on to examine what evidence is required to justify using an alternative to NICE's preferred measure, the EQ-5D.

The key components of the NICE reference case are as follows. The QALY is the recommended measure of the benefits of an intervention. It combines the outcomes of survival and health related quality of life, by placing the latter on a scale where zero represents being dead and one is full health. There are a variety of approaches and techniques for valuing states of health-related quality of life (HRQL). The NICE reference case prefers the generic EQ-5D in adults as reported by the patient or their close carer when they are unable to do so. The EQ-5D comes with a pre-existing value set obtained from a representative sample of the UK general population using the time trade-off technique. Other generic preference-based measures are available, such as HUI3 and SF-6D and these can be used in sensitivity analyses. For children NICE has been less prescriptive, but would like to see standardized and validated preference-based measures to be more widely used.

The EQ-5D can be collected in trials, though for many states it may be more appropriate to collect the data in observational or routine data sets, or to use existing estimates from the literature. Where relevant EQ-5D data are not available, then another solution would be to map from another measure of HRQL or disease severity that has been used in relevant studies and to predict EQ-5D responses from statistical mapping functions. These can be estimated from other data sets containing both instruments. This strategy is accepted by NICE in the absence of EQ-5D data, but it is always second best to the direct use of EQ-5D and may come with a penalty of increased uncertainty.

In some situations, NICE recognises that the EQ-5D may not be appropriate. However, it is difficult to prove a measure of HRQL is or is not valid in a particular patient group in the absence of a gold standard measure. NICE requires empirical evidence to demonstrate the EQ-5D is inappropriate in terms of the properties of content validity, construct validity, and responsiveness. Content validity is concerned with whether the instrument covers all the dimensions of HRQL of importance to patients. Construct validity requires quantitative evidence on whether the measure reflects known differences between groups or converges

with other relevant measures. Responsiveness is the extent to which the EQ-5D reflects changes in HRQL overtime. These criteria would preferably be assessed across the five dimensions of the measure as well as the overall index, though this is rarely done. Careful consideration must be given to the relevance of the variables used to test validity. There will always remain a degree of judgment to be exercised by NICE in any patient group.

The rest of this TSD series considers in more detail the key issues around the use of values in the literature (TSD 9)¹, mapping (TSD 10)² and the alternative methods available where EQ-5D is shown to be inappropriate (TSD 11).³ Finally, the last in the series considers the use of HSUVs in economic models (TSD 12).⁴

The other TSDs in this series offer lists of recommended research in their specific areas and these are not repeated here. In terms of what is discussed in this TSD, the main research gaps are in determining the appropriateness of EQ-5D in different patient groups and further work is required to establish the most appropriate preference-based measure in children.

CONTENTS

1. INTRODUCTION AND BACKGROUND	7
1.1. PURPOSE OF THIS TECHNICAL SUPPORT DOCUMENT	7
2. THE NICE REFERENCE CASE	7
2.1. HEALTH OUTCOMES SHOULD BE EXPRESSED USING QALYS	8
2.2. MEASURING AND VALUING HEALTH.....	10
2.3. WHAT IF PATIENTS ARE UNABLE TO REPORT THEIR HRQL?	12
2.4. WHAT ABOUT THE HEALTH OF CARE-GIVERS?.....	13
2.5. DATA FROM OTHER PREFERENCE-BASED MEASURES OF HEALTH INCLUDED IN CLINICAL TRIALS.....	14
2.6. MEASURING HEALTH-RELATED UTILITY IN CHILDREN.....	14
2.7. IMPLICATIONS FOR DATA COLLECTION AND ANALYSIS.....	15
3. ASSESSING THE APPROPRIATENESS OF EQ-5D	17
3.1. PRACTICALITY.....	18
3.2. RELIABILITY.....	19
3.3. VALIDITY	19
3.3.1 Content Validity	20
3.3.2 Construct validity	20
3.4. RESPONSIVENESS.....	21
3.5. EVIDENCE ON THE APPROPRIATENESS OF EQ-5D	22
4. SUMMARY AND CONCLUSIONS	23
5. SUMMARY OF RECOMMENDATIONS	24
6. REFERENCES	26

1. INTRODUCTION AND BACKGROUND

1.1.PURPOSE OF THIS TECHNICAL SUPPORT DOCUMENT

The Guide to the Methods of Technology Assessment (Methods Guide) describes key aspects of analyses submitted to the technology appraisals programme.⁵ 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 NICE reference case for the measurement and valuation of health for use in cost effectiveness models in health technology assessments (HTA) submitted to NICE. It provides an overview of the NICE methods and introduces the background and rationale to the main topics addressed in the rest of the series. This TSD then goes on to provide guidance on the empirical assessment of the appropriateness of the EQ-5D in different patient groups for NICE submissions. The remaining TSDs in this series go into four issues in greater depth: 1. Identification, reviewing and synthesis of utility data from existing literature; 2. Use of ‘mapping’ to obtain utility values; 3. Alternative methods for generating utility values; and 4. Use of health state utility values in economic models.

2. THE NICE REFERENCE CASE

NICE has outlined a standard approach to the assessment and appraisal of health technologies within its Methods Guide.⁵ The NICE reference case is a core set of methods that should be included in a base case analysis in NICE submissions. The reference case has been developed with regard to promoting consistency in decision-making between appraisals and the principles of transparency, inclusivity, timeliness and methodological robustness.⁶ The measurement and valuation of health is an area of health technology assessment in which there are several controversial areas, particularly because health related quality of life (HRQL) is a concept for which there is no single or objective measure. The NICE Guide to the Methods of Technology Appraisal recognises that a ‘one size fits all’ approach may not always be the most helpful and therefore guidance is also provided on approaches to analysis when the ‘first best’ solution doesn’t seem to work for a given technology. This TSD explains

the recommendations made in the Methods Guide relating to the measurement and valuation of health and provides references to other TSD in this series.

The core components of the NICE reference cover the following issues: the measure of value (i.e. QALYs), the measure of HRQL in adults, who should report HRQL, the inclusion of carers, the role of other measures and the measure of HRQL in children. The next section considers these aspects of the reference case in more detail and their implications for the collection and analysis of HRQL data.

2.1.HEALTH OUTCOMES SHOULD BE EXPRESSED USING QALYS

“For the reference case, cost-effectiveness (specifically cost–utility) analysis is the preferred form of economic evaluation. This seeks to establish whether differences in costs between options can be justified in terms of changes in health effects. Health effects should be expressed in terms of QALYs.” (NICE Methods Guide Section 5.2.11)⁵

Quality adjusted life years (QALYs) are the recommended measure of health outcomes for economic evaluations submitted to NICE. A QALY combines data on length of life with data on the quality of that life. The quality of life component of the QALY represents the value placed on different levels of health or quality of life. This component is anchored at 0, which represents a state as bad as being dead, and 1, which represents full health. Scores of less than zero are theoretically possible and sometimes observed; these represent levels of health that are considered to be worse than being dead. Thus one QALY represents a year in full health. Conventionally in health economic evaluation, QALYs can be summed over time or across individuals. For example, if the prognosis for a patient is expected to be 8 years in full health followed by 2 years in a reduced health state valued at 0.5 (or 50% of full health), their expected QALYs are 9 QALYS: $(8 \text{ years} \times 1.0) + (2 \text{ years} \times 0.5) = 9 \text{ QALYS}$. Also, imagine a treatment available for a specific patient population ($N=10$). If as a result of treatment, the prognosis for 5 patients is 10 years in full health ($5 \times 10 \times 1.0 = 50 \text{ QALYS}$) and 2 years in 50% of full health for the remaining five patients ($5 \times 2 \times 0.5 = 5 \text{ QALYS}$), the total expected QALYs for this patient group is 55 QALYs (or an expected mean of 5.5 QALYs per patient).

Advantages of QALYs include the ability to incorporate a measure of quality of life as well as a measure of longevity, and they can account for the impact of treatment on general health rather than focussing on specific symptoms associated with the condition. In addition,

because the measures of health (in terms of quality and length of life) that make up the QALY are not specific to any given disease, they can be used to compare the amount of benefit derived from different treatments in different conditions.

The NICE Methods Guide states that reference case analyses should always include QALYs as the measure of outcome. However it also refers to the use of alternative measures of health outcomes where there are concerns about the QALY.

“...If the assumptions underlying QALYs (...) are considered inappropriate in a particular case, then evidence to this effect should be produced and analyses using alternative measures may be presented as an additional non-reference case analysis.” (NICE Methods Guide Section 5.2.12)⁵

The key assumptions underpinning the QALY are that: the value placed on the health state (the ‘Q’ element of the QALY) is independent of the duration of that health state; the values placed on health states are independent of when they occur within a profile of health and the sequence in which they occur; and they also usually require risk-neutrality.^{7,8} The use of QALYs is not without debate in the health economics literature, and several studies have reported results of empirical tests of QALY assumptions (for a review see Tsuchiya and Dolan, 2005).⁹ In the context of NICE Technology Appraisals, the focus should be on why the assumptions are less appropriate for a particular condition or treatment.

The NICE Methods Guide does not specify what type of analyses should be presented when the QALY assumptions are shown not to hold. Furthermore, the interpretation of most non-QALY based analyses will be difficult, given that the NICE threshold range is based on a cost per QALY framework and it is not clear what the equivalent value for this range would be if based on alternative measures. Based on a recent review of NICE appraisals, only a handful of economic evaluations were submitted to NICE using non-QALY based analyses.¹⁰ In these cases either health outcomes were excluded altogether (cost-minimisation analyses were presented) or outcomes were expressed as life-years gained, and concerns about the assumptions underlying the QALY were not provided as the rationale for submitting the alternative analyses. A discussion of whether cost-minimisation analysis can ever be appropriate is beyond the remit of this TSD. However, the use of life years gained to represent health outcomes is problematic as it does not reflect impacts on HRQL (positive or negative) and suffers from a lack of comparability to the NICE cost-per-QALY threshold range. Other measures that could be presented as alternatives to the QALY include the

Healthy Year Equivalent (HYE),¹¹ the Saved Young Life Equivalent (SAVE)¹² or contingent valuation. However, HYE and SAVE have not been routinely used for health care policy making and the concerns around them are well documented.¹³ ‘Willingness to pay’ and other contingent valuation methods are consistent with Treasury Guidance for public sector economic evaluations¹⁴ and are widely used in other areas of public policy making, but have been less frequently used for valuing health outcomes for inclusion in economic evaluations. Overall it is expected that the circumstances in which additional alternative analyses would include health outcomes expressed using HYE, SAVE or monetary measures will be rare.

Implications:

- Use QALYs in the base case analysis
- If there are concerns that the assumptions underpinning QALYs do not hold for this specific disease, provide supporting evidence and present alternative analyses.

2.2. MEASURING AND VALUING HEALTH

“...For the reference case, the measurement of changes in HRQL should be reported directly from patients and the value of changes in patients’ HRQL (that is, utilities) should be based on public preferences using a choice-based method. The EQ-5D is the preferred measure of HRQL in adults. The methods to elicit EQ-5D utility values should be fully described. When EQ-5D data are not available or are inappropriate for the condition or effects of treatment, the valuation methods should be fully described and comparable to those used for the EQ-5D...” (NICE Guide to the methods of technology appraisal. Section 5.4)⁵

There are several methods available for measuring the ‘quality’ component of the QALY. NICE has chosen methods where patients report their own HRQL using health status instruments or questionnaires (sometimes referred to as multi-attribute health status classification systems), and then values from existing datasets available for each of the instruments are applied to each description of health (or health state) reported by the patients (sometimes known in the literature as indirect methods). There are several generic and validated instruments available for applying this approach: including the EQ-5D,¹⁵ Short-form (SF)-6D,¹⁶ Health Utilities Index (HUI),¹⁷ Quality of Well-Being (QWB),¹⁸ 15D¹⁹ and

Assessment of Quality of Life (AQoL).²⁰ All have value sets obtained from general population studies that can be applied to the health state described by the classification system. Alternative approaches include patients valuing their own health, usually at repeated points in time to capture the value of treatment (sometimes referred to as ‘direct’ valuation), or developing bespoke descriptions of health states for people to value.

There has been much debate around which approach is best for measuring patient outcomes for use in economic evaluation and decision-making. These well documented debates include whether the focus should be on health-related outcomes or broader measure, whose values should be used to value health states (e.g. patient versus general population), which method of valuation is best (e.g. TTO vs. SG vs. VAS) and which health status instrument should be preferred (e.g. EQ-5D, HUI3 or SF-6D).^{7,13,21} In its Guide to the Methods of Technology Appraisal,⁵ NICE recommends a set of preferred methods which enable it to promote consistency between appraisals and give reassurance to its stakeholders of how certain methods are likely to be viewed by its Appraisal Committees.

NICE recommends an approach that distinguishes between the reporting and valuation of health. It recommends that health status should usually be reported directly by the patients experiencing the condition and/or treatment. It should not be reported by health professionals or by researchers through the development of ‘vignettes’ describing the expected outcomes or experiences of patients. Recommended approaches for when patients are unable to report their own health are described in Section 2.3 below. The guidance also recommends that the values placed on those patient-reported outcomes come from a representative sample of the general population.

Of the main generic preference-based measures noted above, value sets based on UK general population values are available for EQ-5D, SF-6D and HUI 2 (developed for use in children).^{16,17,22} However, it has been shown that the different classification systems produce very different health state values.²³⁻²⁵ This is problematic if comparisons need to be made between evaluations, or with a common reference standard (such as the NICE threshold range). Based on a need for consistency across appraisals, NICE has stated a preference for one of the instruments, the EQ-5D, for use in its economic evaluations. Although this does not imply that EQ-5D should be used in every evaluation (see below).

Implications:

- HRQL should usually be reported by patients using the EQ-5D
- The values placed on changes in health should come from the UK general population using a choice based method
- EQ-5D should be considered the preferred measure for generating health state values in most circumstances

2.3. WHAT IF PATIENTS ARE UNABLE TO REPORT THEIR HRQL?

The Methods Guide emphasises that the health status for a given condition or treatment should be reported directly by patients. However, there may be some circumstances in which this is not possible due to the nature of the condition experienced by the patient (for example, if the condition affects cognitive functioning). The use of a proxy is recommended for these cases where the patient is unable to directly report changes in their own health (NICE Methods Guide Section 5.4.3)⁵ and a version of the EQ-5D for completion by patient proxies has been developed.

The need for a proxy to report health status will not always be clear-cut as the lucidity of respondents is not always easy to judge. Where proxy data are being collected for *de novo* studies a description of the criteria or process used to decide the need for a proxy report should be recorded and reported. Also, it is difficult to determine the most appropriate person to act as proxy for the very reason that a proxy is required: the person is unable to communicate their opinion on their health status. The NICE Methods Guide recommends that where proxies are required, information on the health status of patients is provided by their close carers as the best proxy. Whilst the use of proxy reporting may be inevitable for some patients, the evidence to support agreement between proxy-patient is mixed. There is evidence from EQ-5D and proxy other generic instruments to suggest that agreement between proxies and patients is good for dimensions of health that are more easily observable (e.g. mobility and self-care) but weaker or poor for more subjective dimensions of health (e.g. pain and depression).²⁶⁻²⁸ There is also some evidence showing poor general agreement between patients and proxies.^{29,30} Further research could help determine the best method and person for obtaining proxy reports of health status and whether this varies according to condition.

Implications:

- Where patients are unable to report their health status, this information should come from the patients' close carers

2.4. WHAT ABOUT THE HEALTH OF CARE-GIVERS?

As the population ages, more and more people are providing informal care for close relatives and loved ones, and this can impact on the HRQL of the caregiver. The NICE Reference Case allows for the health impacts on caregivers to be included in the main analysis where treatment is expected to have an effect. It is left to the researcher to decide when there may be such an impact, rather than routinely requiring this information for all appraisals in order to prevent data collection that is likely to be of very limited value. However, it leaves open the question of what constitutes a caregiver. One would not expect professional caregivers to be included in the evaluation, as (compensation for) their time and effort should have been included through staff costs in the evaluation. It is anticipated that the health impact on carers arising from providing informal care to the patient (including the impact on mental health), rather than a general 'family effect' of having emotional connections to someone who is ill, although this is subject to debate in the literature.³¹

It is recommended that the methods used to measure HRQL for the caregiver is the same as that used for the patients. This implies the use of the EQ-5D with the UK population tariff, if EQ-5D is considered to be appropriate for the condition of interest. Where the impact on HRQL of caregivers is included, changes in HRQL over the full period of evaluation should be included. For example, the HRQL of a caregiver may decline as the severity of the patients' condition worsens, but then improve or return to an original level with an increase in supportive care. In these cases the full profile of HRQL change should be included. The QALYs of the caregivers can then be estimated and combined additively with the patient QALYs in the reference case analysis. However both should also be presented separately for purposes of transparency.

Implications:

- Include the impact of treatment on carers health using EQ-5D where this is expected to be important.

2.5. DATA FROM OTHER PREFERENCE-BASED MEASURES OF HEALTH INCLUDED IN CLINICAL TRIALS

Preference-based measures of health-related quality of life other than the EQ-5D (such as the SF-6D or HUI) may have been included in clinical studies used to inform the main estimates of effectiveness in the HTA. In these cases, it is recommended that a sensitivity analysis should be reported using the alternative set of data. Where the analysis is based on a single trial, it is recommended that the data for each treatment arm are used separately, adjusting for baseline differences where necessary. This is particularly important when the health technology assessment is based on a single or a small number of studies as is often the case in Single Technology Appraisals. Other preference-based measures may be also important where EQ-5D has been shown to be inappropriate (see next section for when this may arise).

Implications:

- Present a sensitivity analysis including preference-based measures other than EQ-5D, if they have been included in the clinical trial/s used to inform the effectiveness estimates

2.6. MEASURING HEALTH-RELATED UTILITY IN CHILDREN

The descriptions of health included in many of the health status classification systems are not relevant to young children, including the standard version of the EQ-5D which was originally developed for use in adults. Therefore it is recommended that consideration should be given to using a measure that has been specifically developed for children. The Guide to the Methods of Technology Appraisal⁵ provides the example of the Health Utilities Index version 2 (HUI2), which has been developed specifically for use in children and a value set has been developed based on general population values in the UK.³² The HUI 2 is suitable for self-completion in children aged 8 years and over, or as a proxy version for children aged 5 years and over. Since the publication of the NICE Guide, two other instruments have been developed: the EQ-5D-Youth (Y) and the Child Health Utility 9 dimension (CHU-9D).^{33,34} The EQ-5D-Y is a version of the EQ-5D that has been modified for use in children to make the labels and descriptions of the health domains to be more relevant and accessible to children and young people. It has been developed and validated for children aged 7 to 12

years. The standard adult version of the EQ-5D is considered appropriate for most children aged 12 and upwards. The CHU-9D has been developed specifically for use in children. It has been and validated in both healthy children and children as patients with an age range of 7 to 11 years.

The valuation of health of children raises interesting challenges around the appropriate perspective for valuation. General population values are preferred by NICE, however there are several alternative perspectives that could be taken for the valuation of children's health states that are within the definition of a 'general population perspective'. These include adults valuing the health states as if experienced themselves, adults valuing health states as if experienced by children, or getting values of the health state descriptions from children. Further empirical research would be useful to assess whether these perspectives produce different valuations. A specific value set for the EQ-5D-Y was not available at the time of writing. A value set obtained from members of the general population using a choice-based method (the standard gamble method) is available for the CHU-9D. These values are based on adults valuing the CHU-9D health states as if they experienced the states themselves.³⁵

Implications:

- consider applying an instrument developed for use in children when measuring the health state values of young children.

2.7. IMPLICATIONS FOR DATA COLLECTION AND ANALYSIS

The options for collecting HSUV data are to collect EQ-5D in key clinical trials or observational studies, or to use existing data reported in the literature. Where EQ-5D have not been collected and are not available from other sources, mapping techniques can be used to predict EQ-5D values. Where EQ-5D has been shown to be inappropriate alternative instruments for obtaining HSUVs can be used. These alternative options are reviewed in this section and the links made to the rest of this TSD series.

Source: collecting EQ-5D

An appropriate source for the data on HSUVs may be the main clinical trial/s used to inform the data on effectiveness. This enables the trial data to be used directly within the analysis of HRQL, eliminates concerns about the applicability of the health data to populations from

which the effectiveness estimates are obtained and it enables all the effects of treatment to be included directly in the estimate, including any side-effects of treatment, without the need for adjustment. However, there may be concerns about the generalisability of effectiveness and/or HRQL data to the target NHS population. There may be other circumstances where HSU data are not best collected within the clinical trials, for example if adverse events related to the condition or treatment is rare or not likely to be captured in the trials, where the outcomes of interest are too long term for a typical trial or where the trial does not reflect NHS practice. In these circumstances observational studies may be more appropriate for capturing the impact of the event on HRQL. Such data can then be synthesised with values for other health outcomes or health states, including data from the clinical trials, as needed.

Source: mapping

Where data are not available consideration should be given to ‘mapping’ to EQ-5D from another measure or to using data from the published literature. Mapping uses data from an external source to establish a relationship between the EQ-5D and one or more condition-specific or clinical measure/s that have been included in the main source of effectiveness data; the results from this exercise can then be applied back to the main source of effectiveness used to populate the cost effectiveness model. This is acceptable to NICE as an alternative within the reference case. As this approach provides a method of linking the EQ-5D data from the external source to the effectiveness data and enables multiple criteria to be used to differentiate between EQ-5D values, in many cases it may be preferable to simply obtaining EQ-5D values for a condition from the literature. Mapping can be used to predict EQ-5D directly or can be used to predict the health state descriptions to which values can be assigned in the standard way. However there may also be cases where sufficient EQ-5D data are available in the existing literature that are generalisable to the NICE decision problem. Furthermore, mapping is usually second best to using EQ-5D data collected directly from patients experiencing the treatment of interest. It has been shown that mapping functions tend to overestimate the HSUVs associated with severe health states and underestimate the HSUVs associated with good health, which can misrepresent treatment effects when applied in economic evaluation.³⁶ Further guidance on mapping is provided in TSD 10.²

Source: literature

When identifying health state utility values from the existing literature for NICE assessments, it is not necessary to conduct a systematic review using the full review methods advocated for

identifying the clinical effectiveness data (e.g. using the recommended methods outlined by CRD for effectiveness³⁷). However, care should be taken to ensure the methods of identification and selection are systematic and transparent and that it is made clear why the specific source of data has been selected for use in the analysis. Further guidance on the selection of health-related utility data from the literature is available in TSD 9.¹

Source: alternatives to EQ-5D

Finally, there may be some cases where EQ-5D is considered inappropriate for a specific condition or treatment. The next section focuses on how to determine whether EQ-5D is appropriate for a given condition or treatment. The alternatives to EQ-5D include the use of other preference-based generic measures, condition specific preference-based measures, vignettes and patients own health state valuations. In general, alternative measures based on self-reported health from patients are preferred to vignettes, descriptions based on HRQL are preferred to symptoms, and on the valuation side, values obtained from the general population are preferred to patients and those obtained using choice-based methods are preferred to those that are not (e.g. VAS). Further guidance on these alternative methods for when EQ-5D is considered inappropriate is provided in TSD 11.³

Analysis

When using data from the literature or other secondary sources, it may be necessary to adjust the data in order to make it relevant to the model and the decision problem faced by NICE. For example, the data may reflect the health-related utility associated with a specific condition, but may not include the disutility associated with specific adverse events related to the condition to treatment. There are also issues around how to incorporate uncertainty in the HSUVs estimates. Further guidance on recommended methods for incorporating health state utility data into economics is provided in TSD 12.⁴

3. ASSESSING THE APPROPRIATENESS OF EQ-5D

A crucial decision in the population of an economic model for submission to NICE is whether the EQ-5D is appropriate in the patient group. NICE chose the EQ-5D because it is the most widely used preference-based measure that meets the reference case: a generic measure of HRQL that is self reported and valued using a choice based elicitation technique (TTO) by a representative sample of the UK population. It is believed to be appropriate for

most patient groups, but NICE acknowledges in the Methods Guide that EQ-5D may not be appropriate in all cases. The large scale use of other measures, even other generic preference-based measure of health, would reduce comparability between assessments. NICE has therefore built in a stringent requirement that inappropriateness be empirically demonstrated in situations where other data are submitted. NICE states in its guidance that *'If the EQ-5D is considered inappropriate, empirical evidence should be provided on why the properties of the EQ-5D are not suitable for the particular patient populations. These properties may include the content validity, construct validity, responsiveness and reliability of EQ-5D'*.⁵ Criteria for assessing a measure of health usually include practicality as well as reliability, validity and responsiveness and these are summarised in Table 1.²⁸⁻³⁰

Table 1: A checklist for assessing the inappropriateness of EQ-5D

	Components
Validity	<ul style="list-style-type: none"> • content validity: <ul style="list-style-type: none"> - does the instrument exclude dimensions of health important to patients? - do items appear insensitive? • face validity: <ul style="list-style-type: none"> - are the items relevant and appropriate for the population? • construct validity: <ul style="list-style-type: none"> - does the EQ-5D fail to reflect known differences between groups (preferably across the relevant dimensions) - does EQ-5D and its relevant dimensions fail to converge with other measures of same concept?
Responsiveness	<ul style="list-style-type: none"> • does the instrument fail to reflect known changes in health?

3.1.PRACTICALITY

The *practicality* of an instrument depends on its acceptability to respondents. This can be assessed in terms of how long the instrument takes to administer and the proportion of completed questionnaires. Assignment to the EQ-5D classification requires a short single page instrument and the additional burden arising from its use is minimal, so NICE has not identified this as an argument for regarding it as inappropriate. However, there may be concerns in certain populations with whether it is possible for patients to meaningfully respond, such as when they are extremely ill or cognitively impaired (e.g. case of dementia). In these cases, NICE would accept proxy responses to the EQ-5D, preferably from carers who best know the health of the patient.

3.2.RELIABILITY

Reliability is the ability of a measure to reproduce the same value on two separate administrations when there has been no change in health. This can be over time, between methods of administration or between raters. Evidence on the reliability of the EQ-5D indicates significant random variation between assessments, as is commonly the case with patient reported measures of health.³⁸ This has important implications for the sample size of any study being powered to show differences in EQ-5D. Of more relevance to NICE submissions, it also has implications for the precision of mean health state values used in cost effectiveness models that needs to be fully reflected in the Probabilistic Sensitivity Analysis.

Variation by method of administration may introduce another threat to comparability. Unpublished evidence from the EQ Group suggests there is little or no difference between self report by pencil and paper completion and computer administration.³⁹

3.3.VALIDITY

The assessment of *validity* is more difficult. The lack of a gold standard in the field of health measurement and valuation has resulted in some health economists being rather sceptical about the value of trying to prove validity. Thus a comment by Williams suggested that ‘.....searching for ‘*validity*’ in this field,, is like chasing will o’ the wisp, and probably equally unproductive’.⁴⁰ The challenge of assessing the validity of such data pervades the measurement of all psychological phenomena and has been met in the psychometric literature by the development of various tests that can be adapted with care for use on the EQ-5D.

The EQ-5D is composed of a descriptive system and a set of values. The descriptive system of the EQ-5D has five dimensions (mobility, self care, usual activities, pain/discomfort and anxiety/depression) and in the reference case each dimension has 3 levels of no problem, some problems and severe problems. The EQ-5D should be used with a pre-existing ‘value set’ obtained from a representative sample of the UK general population using a preference elicitation technique known as time trade-off.²² The appropriateness of the EQ-5D is concerned with the descriptive system rather than the values per se, since this is the part that defines the coverage and sensitivity of the instrument, though most evidence is concerned with the index. The appropriateness of the descriptive system can be assessed in terms of the content and face validity, construct validity and responsiveness.⁴¹

3.3.1 Content Validity

In terms of content, the most fundamental question concerns the definition of the construct of HRQL. NICE does not provide a definition of HRQL in the Methods Guide for Technology Appraisal nor is there a single definition agreed by researchers in the field. The definition of health provided in the constitution of the World Health organisation of ‘*A state of complete physical, mental and social well-being*’⁴² has been very influential in the field. A more recent definition of HRQL provided by researchers that represents what most people in the field mean by HRQL is: ‘*A person’s subjective perception of the impact of health status, including disease and treatment, on physical, psychological and social functioning and well-being*’.⁴³ EQ-5D can be seen as a combination of physical functioning (mobility, self care), psychological health (depression and anxiety), social functioning (that may be included in usual activities) and symptoms (pain and discomfort). Within this conceptual framework, the precise content can be assessed using qualitative methods such as depth interviews and focus groups of patient and this is the approach recommended by the US Food and Drug Administration (FDA) for patient reported outcome measures to support labelling claims.¹³ The face validity of a descriptive system can be assessed using cognitive interview techniques to establish whether patients, for example, understand the descriptions in the way they are intended to be understood.

Qualitative techniques may be helpful in making a case against the EQ-5D, but they are not sufficient. Simply showing that a particular concern is not explicitly in the descriptive system of EQ-5D does not imply that the impact on HRQL is not appropriately reflected in one or more of the 5 dimensions in the instrument.

3.3.2 Construct validity

Construct validity is the extent to which the scores produced by a measure agrees with other measures or indicators of the dimensions of HRQL considered relevant to the patient group (such as those identified by qualitative work). There are two commonly used tests in the psychometric literature for examining the construct validity. One approach is to examine whether it is able to differentiate between groups thought to differ in terms of their health (i.e. known group differences), and the other is the extent to which it correlates with another measure of health (i.e. convergent validity). Construct validation is best performed at the dimension level, but evidence is often only available on the index that is a preference weighted aggregate across the dimension.⁴⁴ These tests can never prove the validity of an instrument, but they provide evidence on the degree to which a measure is valid at measuring

the concept being tested. For tests based on known group differences this depends on the basis for the groupings used to assess known group differences. Where these are other self-report instruments of dimensions of interest, such as other scales of mobility or self care, then these can be useful in assessing whether the EQ-5D descriptive system is sensitive to differences. However, evidence in the psychometrics literature often uses clinical measures such as visual acuity, respiratory function or symptoms of schizophrenia that may have only a weak relationship to HRQL in any case. Great care must be taken to scrutinise the measures being used to establish known group differences or convergence and to establish that these are themselves appropriate.

Convergent validity can be assessed by the strength of association between EQ-5D and another measure of preferences, such as another generic preference-based measure (e.g. HUI3 and SF-6D). A lack of association with other measures, or the EQ-5D failing to reflect differences between groups found by HUI3 or SF-6D may suggest the EQ-5D is not valid in this group. Another potential source of evidence comes from convergence with directly administered TTO, SG or VAS, but this test is limited by the fact that they reflect patient values rather than those of the general population.

3.4. RESPONSIVENESS

A related empirical test is *responsiveness*, which is the ability of an instrument to measure “clinically significant changes” in health. For populating cost effectiveness models, this is less important than validity since models requires mean health state values. However, it does provide another source of data on whether EQ-5D is able to reflect known differences over time. Psychometric tests of responsiveness suffer from the same limitation as construct validation, since the alleged change in health comes from clinical assessment or before and after an intervention that may not have improved patient health. Furthermore, the psychometric literature uses measures of responsiveness such as the ‘effect size’, where the mean change in score is divided by either the standard deviation at baseline or the standard deviation of the change.⁴⁵ A common assumption in the psychometric literature is that for a given health change, the measure with the larger effect size is the better measure.^{46,47} Where the objective is to minimise the sample size, this makes sense. However, when the purpose is to compare the size of change between treatments as part of an economic evaluation, within or between conditions, it is the value of the change which matters.⁴⁸ Effect sizes do not indicate the value or importance of a change.

Finally a specific preference study may be undertaken to establish whether a dimension of HRQL excluded from EQ-5D could be important. This could take the form of an ‘add-on’ study, where it is established whether the supposed missing dimension is valued over and above the existing dimensions of the EQ-5D. The addition of a sleep dimension, for example, did not significantly change the values given to EQ-5D health states.⁴⁹ This does not mean that sleep is unimportant, simply that its impact may be indirectly felt through one of the existing five dimensions. Further work of this kind is currently ongoing.

3.5. EVIDENCE ON THE APPROPRIATENESS OF EQ-5D

An empirical literature on this subject has only recently begun to emerge. Most evidence tends to be around the aggregated index rather than the individual level dimensions and to be quantitative rather than qualitative. It uses a mix of other generic preference-based measures, directly elicited preferences (using TTO and SG), clinical measures and assessments before and after interventions. Much of this testing does not meet the standards outlined above. The studies providing the evidence in most cases were not designed to examine validity or responsiveness, and so tend to be under-powered for this purpose and use variables to define groups or changes over time that may not reflect differences important to patients or that would be valued by the general public.

This TSD is not able provide a detailed review of the appropriateness of EQ-5D across patient groups. The literature is too large and it is growing all the time. Evidence from recent reviews suggests the EQ-5D is probably not appropriate for assessing the impact hearing loss,⁵⁰ some specific forms of visual impairment⁵¹ and schizophrenia.⁵² Unfortunately the evidence to support alternative generic preference-based measures in these populations is also currently limited. However, it would seem that EQ-5D is appropriate in areas including depression and anxiety,⁵³ number of key cancers,⁵⁴ cardiovascular disease,⁵⁵ type 2 diabetes,⁵⁶ asthma and chronic obstructive pulmonary disease.⁵⁷ The reader is recommended to consult TSD 9¹ on methods for identifying the relevant literature where reviews have not been undertaken. Where there is not sufficient evidence in the existing literature, then additional empirical work may be required including qualitative work on the content of the EQ-5D and psychometric evidence on validity and responsiveness.

Implications:

The EQ-5D is assumed to be appropriate unless it is empirically demonstrated not to be the case for a given patient group and its treatment.

Evidence to support this should be based on a systematic review of evidence of EQ-5D data in the specific condition

It should include data on the content validity, construct validity and responsiveness of EQ-5D

Consideration should be given to the quality of the studies, including the sample size and to the validity of the measures used for comparison in the identified studies

4. SUMMARY AND CONCLUSIONS

NICE requires consistency in its decision making and so has developed a reference case of methods for Technology Appraisal that includes the measurement and valuation of health. The key components of this reference case are that QALYs should be the measure of value. The preferred instrument for putting 'Q' into the QALY is the EQ-5D in adults as reported by the patient, or their carer when they are unable to do so, and valued by a general population value set obtained using TTO. Other generic preference-based measures can be used in a sensitivity analyses. For children NICE has been less prescriptive, but would like to see standardized validated preference-based measures to be more widely used.

The EQ-5D can be collected in trials, though for many states it may be more appropriate to collect it in observational or routine data sets, or to use existing estimates from the literature. Where relevant EQ-5D data are not available, then another solution would be to map from another HRQL measure that has been used in a group of patients in the relevant health states, and EQ-5D responses predicted from existing statistical mapping functions. This is always second best to the direct use of EQ-5D and may come with a penalty of increased uncertainty. In some situations, NICE recognises that the EQ-5D may not be appropriate. However, it is difficult to prove this is the case in the absence of a gold standard measure. To do so requires

empirical data on the validity of the content in patients and empirical data on how well it performs in terms of psychometric criteria.

The rest of this TSD series considers in more details the key issues around the use of values in the literature (TSD 9),¹ mapping (TSD 10)² and the alternative method available where EQ-5D is shown to be inappropriate (TSD 11).³ Finally, the last in the series considers the use of HSUVs in economic models (TSD 12).⁴

The other TSDs in this series offer lists of recommended research in their specific areas and these are not repeated here. In terms of what is discussed in this TSD, the main research gaps are in determining the appropriateness of EQ-5D in different patient groups. There is some research currently ongoing looking at this issue, but gaps remain and it will need updating. These will be important resources to make readily available to those looking to submit evidence to NICE. Further work is required on the most appropriate preference-based measure in children.

5. SUMMARY OF RECOMMENDATIONS

Reference case

- The QALY is the measure of the benefit of treatment.
- Patient self report should be used to describe the change in health,
 - although where they are unable to provide this, then the information should come from the patients' close carers.
- The EQ-5D should be used to collect data from patients on their health, and a set of values obtained from the UK general population using the time-trade off method applied to generate health-related utilities.
- Where it is important, the impact of an intervention on carers can be included and measured using the EQ-5D.
- Other preference-based measures of health can be included in sensitivity analysis, if they have been included in the clinical trial/s used to inform the effectiveness estimates.

- Consider using an instrument developed for use in children when obtaining health state utility values.

Implications

- Data on EQ-5D can be collected in clinical trials, observational studies, routine sources or the literature depending on the requirements of the cost effectiveness model.
- Mapping from another HRQL measure offers a solution where relevant EQ-5D is not available, but may come with increased uncertainty.
- The EQ-5D is assumed to be appropriate unless it is empirically demonstrated not to be the case for a given patient group and its treatment.
 - Evidence to support this should be based on a systematic review of evidence of EQ-5D data in the specific condition.
 - It should include data on the content validity, construct validity and responsiveness of EQ-5D.
 - Consideration should be given to the quality of the studies, including sample size and completion rates.
 - Consideration should be given to the aspects of health that are expected to change and to the validity of the measures used for comparison in the identified studies.
- There are alternative methods, but the use of these reduces comparability and so evidence for their use must be convincing.

6. REFERENCES

1. Papaioannou, D., Brazier, J.E., Paisley, S. NICE DSU Technical Support Guide 9: The identification, review and synthesis of health state utility values from the literature. 2011; available from <http://www.nicedsu.org.uk>
2. 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>
3. 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>
4. Ara, R., Wailoo, A.J. NICE DSU Technical Support Document 12: The use of health state utility values in decision models. 2011; available from <http://www.nicedsu.org.uk>
5. National Institute of Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. 2008. London, NICE.
6. Longworth, L., Longson, C. NICE Methodology for Technology Appraisals: cutting edge or tried and trusted? *Pharmacoeconomics* 2008; 26(9):729-732.
7. Brazier, J., Ratcliffe, J., Salomon, J., Tsuchiya, A. Measuring and valuing health benefits for economic evaluation. First edition ed. Oxford University Press, Oxford; 2007.
8. Weinstein, M.C., Torrance, G., McGuire, A. QALYs: the basics. *Value in Health* 2009; 12 Suppl 1:S5-S9.
9. Tsuchiya, A., Dolan, P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Medical Decision Making* 2005; 25(4):460-467.
10. Tosh, J.C., Longworth, L.J., George, E. Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. *Value in Health* 2011; 14(1):102-109.
11. Mehrez, A., Gafni, A. The healthy-years equivalents: how to measure them using the standard gamble approach. *Medical Decision Making* 1991; 11(2):140-146.
12. Nord, E. An alternative to QALYs: the saved young life equivalent (SAVE). *BMJ* 1992; 305(6858):875-877.
13. Drummond, M., Sculpher, M., Torrance, G., O'Brien, B., Stoddart, G. Methods for the economic evaluation of health care programmes. 3rd edition ed. Oxford University Press, Oxford; 2005.
14. The Treasury. The Green Book: appraisal and evaluation in central government. 2003. London, The Stationary Office. 11 A.D.

15. Brooks, R. EuroQol: the current state of play. *Health Policy* 1996; 37(1):53-72.
16. Brazier, J., Usherwood, T., Harper, R., Thomas, K. Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology* 1998; 51(11):1115-1128.
17. Feeny, D., Furlong, W., Boyle, M., Torrance, G.W. Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics* 1995; 7(6):490-502.
18. Kaplan, R.M., Anderson, J.P. A general health policy model: update and applications. *Health Services Research* 1988; 23(2):203-235.
19. Sintonen, H. The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine* 2001; 33(5):328-336.
20. Hawthorne, G., Richardson, J., Osborne, 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.
21. Dolan, P. Output measures and valuation in health. In: Drummond M., McGuire A., eds. *Economic evaluation in health care: merging theory with practice*. First edition ed. Oxford University Press; Oxford: 2001.
22. Dolan, P. Modeling valuations for EuroQol health states. *Medical Care* 1997; 35(11):1095-1108.
23. 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.
24. Longworth, L., Bryan, S. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics* 2003; 12(12):1061-1067.
25. Whitehurst, D.G., Bryan, S. Another Study Showing that Two Preference-Based Measures of Health-Related Quality of Life (EQ-5D and SF-6D) are not Interchangeable. But why Should we Expect Them to be? *Value in Health* 2011.
26. Dorman, P.J., Waddell, F., Slattery, J., Dennis, M., Sandercock, P. Are proxy assessments of health status after stroke with the EuroQol questionnaire feasible, accurate, and unbiased? *Stroke* 1997; 28(10):1883-1887.
27. Hung, M.C., Yan, Y.H., Fan, P.S., Lin, M.S., Chen, C.R., Kuo, L.C. et al. Measurement of quality of life using EQ-5D in patients on prolonged mechanical ventilation: comparison of patients, family caregivers, and nurses. *Quality of Life Research* 2010; 19(5):721-727.
28. Pickard, A.S., Johnson, J.A., Feeny, D.H., Shuaib, A., Carriere, K.C., Nasser, A.M. Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and Health Utilities Index. *Stroke* 2004; 35(2):607-612.

29. Coucill, W., Bryan, S., Bentham, P., Buckley, A., Laight, A. EQ-5D in patients with dementia: an investigation of inter-rater agreement. *Medical Care* 2001; 39(8):760-771.
30. Jonsson, L., Andreasen, N., Kilander, L., Soininen, H., Waldemar, G., Nygaard, H. et al. Patient- and proxy-reported utility in Alzheimer disease using the EuroQoL. *Alzheimer Disease & Associated Disorders* 2006; 20(1):49-55.
31. Bobinac, A., Van Exel, N.J., Rutten, F.F., Brouwer, W.B. Caring for and caring about: disentangling the caregiver effect and the family effect. *Journal of Health Economics* 2010; 29(4):549-556.
32. McCabe, C.J., Stevens, K.J., Brazier, J.E. Utility scores for the Health Utilities Index Mark 2: an empirical assessment of alternative mapping functions. *Medical Care* 2005; 43(6):627-635.
33. Stevens, K. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Quality of Life Research* 2009; 18(8):1105-1113.
34. Wille, N., Badia, X., Bonsel, G., Burstrom, K., Cavrini, G., Devlin, N. et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Quality of Life Research* 2010; 19(6):875-886.
35. Stevens, K. Valuation of the Child Health Utility Index 9D (CHU9D). 2010; available from <http://www.shef.ac.uk/scharr/sections/heds/dps-2010>
36. Rowen, D., Brazier, J., Roberts, J., Rowen, D., Brazier, J., Roberts, J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health & Quality of Life Outcomes* 2009; 7:27.
37. NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness. 2011. York.
38. Streiner, D.L., Norman, G.R. "Precision" and "accuracy": two terms that are neither. *Journal of Clinical Epidemiology* 2006; 59(4):327-330.
39. Lloyd, A.J., Kind, P., Thompson, T., Leese, B., Nixon, A., Quadri, N. Paper to web: equivalence testing of EQ-5D report to the Department of Health. 2011.
40. Williams, A. The measurement and valuation of health: a chronical. 1995. University of York. Discussion paper 136.
41. Streiner, D.L., Norman, G.R. Health measurement scales: a practical guide to their development and use. Oxford University Press, Oxford; 1997.
42. World Health Organisation. Constitution of the World Health Organisation. 1948. Geneva, Switzerland.
43. Leidy, N.K., Revicki, D.A., Geneste, B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. *Value in Health* 1999; 2(2):113-127.

44. Brazier, J., Deverill, M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Economics* 1999; 8(1):41-51.
45. Guyatt, G.H. Making sense of quality-of-life data. *Medical Care* 2000; 38(9 Suppl):II175-II179.
46. Fitzpatrick, R., Ziebland, S., Jenkinson, C., Mowat, A., Mowat, A. A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *Journal of Rheumatology* 1993; 20(3):429-436.
47. Katz, J.N., Phillips, C.B., Fossel, A.H., Liang, M.H. Stability and responsiveness of utility measures. *Medical Care* 1994; 32(2):183-188.
48. Walters, S.J., Brazier, J.E. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research* 2005; 14(6):1523-1532.
49. Yang, Y., Brazier, J., Tsuchiya, A. The effect of adding a sleep dimension to the EQ-5D. Paper Presented to the HESG Meeting, University of East Anglia, January 2008 2008.
50. Barton, G.R., Bankart, J., Davis, A.C., Summerfield, Q.A. Comparing utility scores before and after hearing-aid provision. *Applied Health Economics and Health Policy* 2004; 3(2):103-105.
51. Tosh, J., Brazier, J., Evans, P., Longworth, L. A review of generic preference-based measures of health-related quality of life in visual disorders. *Value in Health* 2011; 15 (1): 118-127.
52. Papaioannou, D., Brazier, J., Parry, G. How valid and responsive are generic health status measures, such as the EQ-5D and SF-36, in schizophrenia? A systematic review. *Value in Health* 2011; 14 (6): 907-920.
53. Papaioannou, D., Peasgood, T., Parry, G., Brazier, J. Are the EQ-5D and SF-36 appropriate in mental health? Euroqol Group Scientific Meeting, Oxford, 2011.
54. Pickard, A.S., Wilke, C.T., Lin, H.W., Lloyd, A. Health utilities using the EQ-5D in studies of cancer. *Pharmacoeconomics* 2007; 25(5):365-384.
55. Dyer, M.T., Goldsmith, K.A., Sharples, L.S., Buxton, M.J. A review of health utilities using the EQ-5D in studies of cardiovascular disease. *Health and Quality of Life Outcomes* 2010; 8:13.
56. Janssen, M.F., Lubetkin, E.I., Sekhobo, J.P., Pickard, A.S. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. *Diabetic Medicine* 2011; 28(4):395-413.
57. Pickard, A.S., Wilke, C., Jung, E., Patel, S., Stavem, K., Lee, T.A. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. *Respiratory Medicine* 2008; 102(4):519-536.