# NICE DSU TECHNICAL SUPPORT DOCUMENT 9: THE IDENTIFICATION, REVIEW AND SYNTHESIS OF HEALTH STATE UTILITY VALUES FROM THE LITERATURE

REPORT BY THE DECISION SUPPORT UNIT

October 2010

Diana Papaioannou, John Brazier, Suzy Paisley

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 would like to thank Jenny Dunn for her help in preparing and formatting the report. We would like to thank Julie Glanville, Colin Green and Chris Hyde who commented on this document. We would also like to thank Helen Starkie and Jennifer Priaulx (NICE) who commented on 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:**

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

# EXECUTIVE SUMMARY

## BACKGROUND

Health state utility values (HSUVs) are important parameters in decision models, and NICE requires evidence that HSUV estimates from the published literature have been identified and selected systematically. NICE provides a reference case analysis with specific requirements for how HSUVs are derived, including a preference for HSUV values derived from the EQ-5D. This Technical Support Document (TSD) discusses how to systematically identify and select HSUVs from the literature in order to meet the requirements of the NICE Methods Guide. In addition, the TSD provides guidance on quality and relevance assessment, data extraction, selection of values and synthesis. Two case studies are used throughout this document; a review of HSUVs in osteoporosis-related conditions and a review of HSUVs in breast cancer.

## SCOPING AN HSUV REVIEW

The aim at the scoping stage is to characterise the precise HSUVs that need to be captured by the review in order to inform the decision model. Two key elements must be defined: 1) the specific health states required for the decision model and 2) the type of HSUV data required by NICE. It is recommended that the scope of the review be kept broad at first since the precise nature of and quantity of available HSUV evidence may not be known at the scoping stage; further refinement of the scope is undertaken during the evidence selection stage. HSUVs may be required for a number of health states, subgroups and over different time periods, and this may extend to states beyond the primary condition explored in the decision model. The type of HSUV data required is preferred by NICE to be estimated from the EQ-5D. However, there are instances when the EQ-5D is considered inappropriate for some patient populations.

## IDENTIFYING AND SELECTING THE EVIDENCE

Searching for HSUV reviews requires a broad electronic database search using an extensive list of search terms for HSUV concepts, as well as scrutinising reference lists of retrieved studies. Ideally, other supplementary search techniques should be used such as contact with experts, citation and author searching. Selecting the evidence for the review involves refining the scope and inclusion/exclusion criteria as the nature of the evidence based is determined.

This requires some preliminary data extraction of three key details: i) details of the population describing the health state (e.g. age, sex, disease severity); ii) details of the approach used to describe the health state and iii) HSUV elicitation technique e.g. TTO, SG, VAS. Based on the findings from this preliminary data extraction, a picture can be built up about the nature and quantity of evidence available, which in turn will allow decisions on studies to include or exclude. Transparent documentation is required for the identifying and selection of evidence.

## QUALITY AND RELEVANCE ASSESSMENT

There are no agreed reporting standards for HSUV studies. Key criteria of quality assessment of HSUV studies proposed are sample size, respondent selection and recruitment, inclusion/exclusion criteria, response rates, numbers lost to follow-up (and reasons), and methods of missing data analysis. As important as quality assessment is the relevance of the data to the decision model and to the agency to which the model will be submitted. For NICE, this involves looking at how well the data matches the NICE reference case analysis.

## DATA EXTRACTION

Data extraction largely follows the same principles as that for clinical effectiveness reviews. Therefore, general information such as author or country of publication; study characteristics such as inclusion/exclusion criteria; participant characteristics such as age, sex, disease characteristics and study setting. Information relating to the instruments used to collect descriptive data, valuation techniques and source of values is important to record, as well as descriptive statistics on the results.

## DATA PRESENTATION

Data presentation involves providing: i) characteristics of included studies, ii) HSUVs used in the decision model including full justification for their use and where synthesised HSUVs used an account of heterogeneity is required, iii) quality and relevance assessment of included studies, iv) modifications to HSUVs for use in the decision model and v) details of sensitivity analyses undertaken.

## DATA SYNTHESIS

Methods for selecting and synthesising depend on the availability of HSUV data. There are four situations to consider: i) one set of relevant HSUVs – there should form the central estimates for the states used in the model; ii) multiple sets of HSUVs meet the criteria of the review -then the selection of values used in the model needs to be justified and alternative and less relevant HSUVs should be used in a sensitivity analysis to better understand the impact of this parameter on the model. Where more than one set of relevant values are sufficiently homogenous (i.e. collected from the same patient population using the same instrument and valued using the same UK value) then pooling should be considered as a way to improve the precision of the estimates of the mean HSUVs and their variances; and iv) no directly relevant values (e.g. such as the reference case has not been used or the patient group is not appropriate for the model) - values still need to be selected and justified and some form of meta regression may help in this situation in order to better understand the causes of variation and hence provide support for the values selected.

## RECOMMENDATIONS FOR UNDERTAKING HSUV REVIEWS

- The scope and identification of the evidence for HSUV reviews need to be kept broad initially
- Ideally use a variety of resources and methods to identify relevant studies e.g. electronic database searching, reference list checking, contact with experts etc.
- The scope of the review and inclusion/exclusion criteria will be refined during the stage of evidence selection according to the nature of the evidence base
- Selecting evidence for HSUV reviews is an iterative process and may involve preliminary data extraction of key characteristics (population details, approach used to describe the health state, elicitation technique). Based on that data, decisions can be made on how to amend and develop inclusion criteria further.
- Selection of included studies must be well justified and explicit. A record of reasons for study inclusion and exclusion (i.e. those studies identified as possible but ultimately excluded) must be kept.
- Criteria for quality and relevance have been suggested in this guide (see Box 3 and 4)
- Where there is more than one set of values meeting the reference case and that are relevant to the model, then the final selection needs to be justified with sensitivity

analyses using alternative values, and consideration given to synthesis to improve precision of estimated HSUVs.

- Data presentation must include an account of the search undertaken to identify studies, characteristics of included studies, HSUVs used in the decision model (with justification), quality and relevance assessment and modifications made to values used in the model.

## CONCLUSION

NICE requires evidence that HSUV estimates from published literature have been identified and selected systematically. The principles of systematic reviewing for clinical effectiveness reviews can inform some aspects of how to identify and select utilities systematically, but there are unique issues to be explored in the scoping and identification of evidence of HSUVs reviews. The process of identifying and selection evidence differs in that often a sequence of searches may be required, rather than one literature search. Study selection also informs the inclusion criteria in terms of refining the type of HSUV data to be included according to the evidence available. The process of evidence assessment involves both quality and relevance assessment. The final selection of values used in the model needs to be justified and sensitivity analysis undertaken of alternative possible values.

# CONTENTS

## TABLES AND FIGURES

**Abbreviations and definitions**

| | |
|---|---|
| Decision model | In the context of economic evaluation, a decision analytic model uses mathematical relationships to define a series of possible consequences that would flow from a set of alternative options being evaluated.[ii] |
| HRQL | Health-related quality of life |
| HSUV/HSUVs | Health state utility value(s) |
| HTA | Health technology assessment |
| Precision | The ability of the literature search to reject irrelevant material. |
| QALY | Quality adjusted life year |
| RCT | Randomised controlled trial |
| Responsiveness | Ty ability of a measure to detect changes over time |
| SG | Standard Gamble |
| Sensitivity | The ability of the literature search to find all relevant material precision is its ability to reject irrelevant material |
| TSD | Technical support document |
| TTO | Time trade-off |
| Validity | The degree to which a measure measures what it claims to measure |
| VAS | Visual analogue scale |

[ii] Briggs et al, Decision Modelling for Health Economic Evaluation, 2006, Oxford.

# 1. INTRODUCTION AND BACKGROUND

## 1.1. PURPOSE OF THIS TECHNICAL SUPPORT DOCUMENT

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

This TSD is concerned with *the identification, review and synthesis of health state utility values from the literature.* Whilst the TSD looks at systematically reviewing HSUVs in general, particular emphasis is placed on the review of HSUVs in order to generate values for the parameters of decision models in health technology assessment (HTA) submissions to NICE.

## 1.2. MEASUREMENT AND VALUATION OF HEALTH: IDENTIFICATION, REVIEW AND SYNTHESIS OF UTILITY VALUES

The NICE Methods Guide requires economic evaluations for HTA submissions to NICE to be submitted as a cost-effectiveness analysis presented in the form of an incremental cost per quality adjusted life years (QALYs) for the appropriate time horizon.[2] HSUVs provide the essential quality weight for calculating the quality-adjusted-life-year (QALY) of an intervention.[3] The Methods Guide also provides details of a reference case analysis in terms of the preferred instrument (EQ-5D) for generating the HSUVs and what to do when EQ-5D data are not available or is regarded as inappropriate.

The NICE Methods Guide states that 'the use of HSUV estimates from the published literature must be supported by evidence that demonstrates they have been identified and selected systematically'[4] (section 5.4, p.38). There is little guidance provided on how to identify HSUV evidence systematically, including the EQ-5D, for the health states used in decision models[iii] to estimate the incremental cost effectiveness. Previously, decision models

[iii] Note that in the vast majority of cases, economic evaluations to NICE will involve a decision model, and this is the terminology used throughout this document. However, in some instances a submission may be based on other economic analyses.

have tended to present a single set of HSUVs to inform such parameters, with little justification as to why they have been selected above other values.[5] Whilst little detail is provided in the NICE Methods Guide as to how to search for and select HSUVs within the literature, methods for identifying evidence for systematic reviews undertaken to generate reliable estimates of clinical effects for use in decision models are well developed.[6,7] With a growing literature of empirically derived HSUVs, it is going to be increasingly important to ensure that the methods used to identify and select HUSVs are systematic and transparent to justify the values are used in decision models.

This TSD will discuss the issues in systematically reviewing HSUVs and where appropriate provide guidance. This will include

- How to identify, and select HSUVs for review from the published literature in a systematic way
- How to systematically review HSUV data in terms of quality and relevance and how this may differ from reviews of clinical effects
- How to select and in some cases synthesise HSUV values across studies for use in a decision model.

This guide will look at each of the processes typically undertaken in systematic reviews of clinical effects and provide guidance on methods for each in the context of systematically reviewing HSUV data.

### *1.2.1. The NICE Methods Guide requirements*

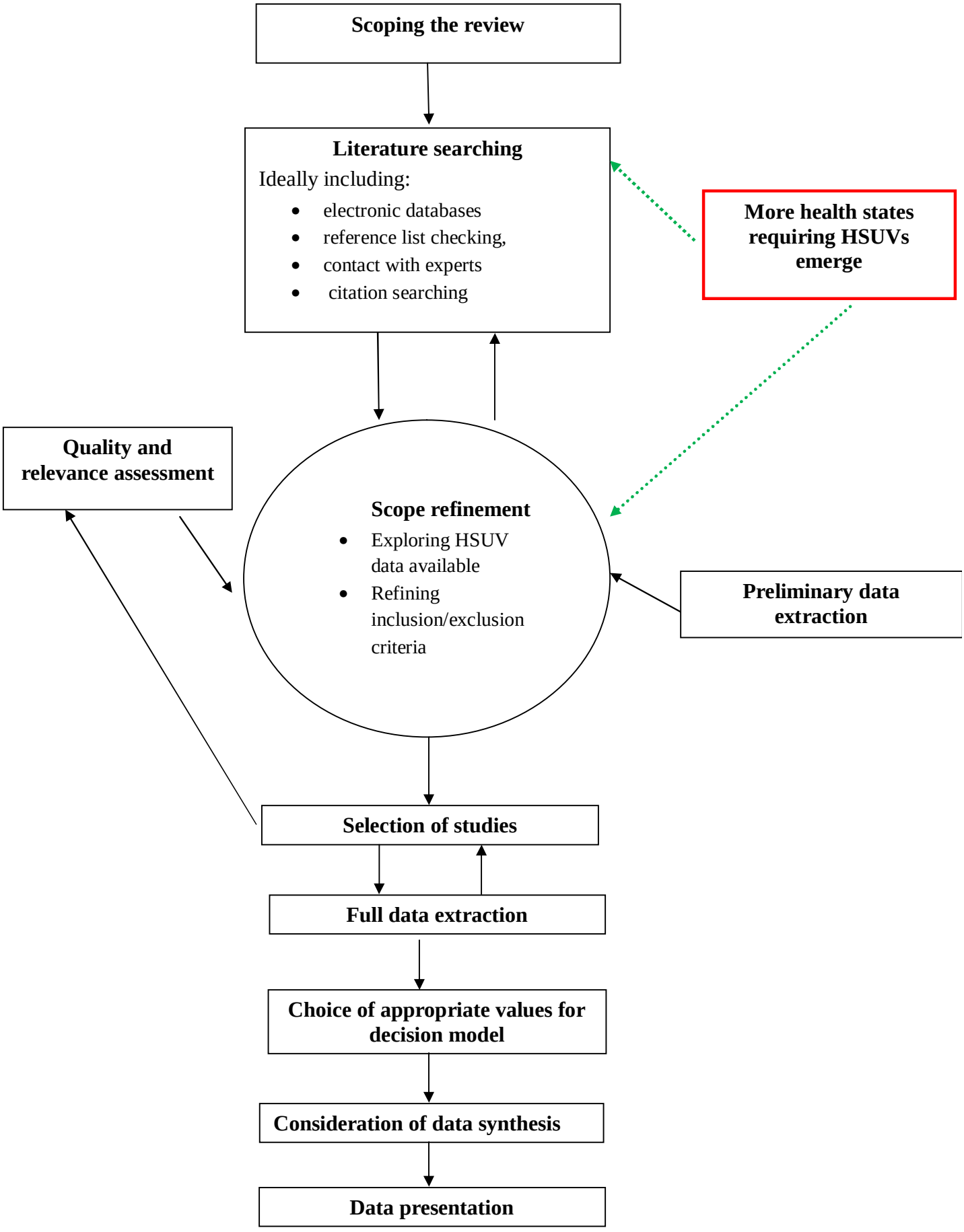
The NICE methods guide requires that HSUVs identified in the published literature are 'identified and selected systematically'[2] (section 5.4, p.38) without necessarily conducting a full systematic review. Within this TSD, Section 2 provides guidance on how to scope systematically and section 3 provides guidance on how to search and select HSUVs systematically from the published literature. In addition, section 7 of this TSD suggest ways in which to present the results of reviewing HSUVs in the literature which is important in order to provide evidence that HSUVs have been systematically identified and selected.

The remaining sections in this TSD provide guidance on the processes involved when undertaking systematic reviews of HSUVs. This includes quality and relevance assessment (section 4), data extraction (section 5) and where appropriate data synthesis (section 6). In

addition, areas of methodological uncertainty are highlighted and research recommendations made (section 8.2).

Whilst processes such as quality and relevance assessment, data extraction and synthesis are not discussed as requirements in the NICE methods guide, they are nevertheless important ways of making the process of reviewing HSUVs more robust, systematic and transparent and can add value to a submission. For instance, undertaking quality and relevance assessment (as discussed in section 4) will provide information that a NICE Appraisal Committee can use to consider the strengths and limitations of the HSUVs used in the decision model. Thus, there are advantages to conducting a systematic review of HSUVs. However, this needs to be weighted against the additional resources required to undertake a systematic review. Figure 1 illustrates the processes covered by this TSD.

**Figure 1: Reviewing HSUVs**



### *1.2.2. Case studies*

Throughout, the guide we refer to two case studies: a systematic review of HSUVs in osteoporosis-related conditions [8] and a systematic review and meta-analysis of HSUVs in breast cancer. [9] The characteristics of these two reviews are broadly described in Table 1.

**Table 1: Key characteristics of case studies used to illustrate this TSD**

| | **Population** | **Health states** | **Types of HSUVs identified** |
|---|---|---|---|
| Peasgood, T et al (2009). An updated systematic review of HSUVs for osteoporosis related conditions. [8]<br><br>Other relevant papers: [10-12] | Male and female adults with conditions related to osteoporosis | • Established osteoporosis<br>• Vertebral fracture<br>• Hip fracture<br>• Wrist fracture<br>• Shoulder fracture | Own health: SG, TTO, VAS<br><br>Preference-based measures EQ-5D, SF-6D, HUI3, QWB<br><br>Bespoke vignettes |
| Peasgood, T et al. A review and meta analysis of HSUVs in breast cancer [9] | Adult population with breast cancer | • Screening-related states<br>• Preventative states<br>• AEs in breast cancer and its treatment<br>• Non-specific breast cancer<br>• MBC states<br>• EBC states | Own health: SG, TTO, VAS<br><br>Preference-based measures EQ-5D, HUI3,<br><br>Bespoke vignettes |

AEs-adverse events, EBC-early breast cancer, MBC- metastatic breast cancer,

SG-standard gamble, TTO-time-tradeoff, VAS-visual analogue scale, QWB- quality for wellbeing

# 2. SCOPING THE REVIEW

The aim at the scoping stage is to characterise the precise HSUVs that need to be captured by the review in order to inform the decision model. Whilst in reviews of clinical effects, methods guides recommend structuring the review question according to the PICO question (Patient, Intervention, Comparison, and Outcome),[6,7] this is not a useful framework for scoping HSUV reviews. Firstly, the 'Intervention' and Comparison' elements in PICO are not usually relevant to HSUVs reviews, where the aim is often to identify HSUV data for particular health states that are not necessarily attached to an intervention. Secondly, decision models typically require a series of HSUVs as they examine the whole treatment pathway and thus what happens to patients over a longer time horizon (e.g. rest of the patient's life). For example over a period of treatment, HSUVs may be required for receiving effective treatment, receiving non-effective treatment, each individual adverse event, disease progression or stable disease. Thirdly, whilst reviews of clinical effects often focus on specific study designs (with evidence from randomised controlled trials (RCT) often being seen as the gold standard by which to assess clinical efficacy),[6,7] HSUV data is not exclusively reported in RCTs. Often HSUVs are reported in observational studies as well as other cost-effectiveness studies such as HTAs and economic evaluations, and thus limiting by study design is not appropriate for reviews of HSUVs.

We recommend when scoping reviews of HSUVs to define two key elements: 1) the specific health states required for the decision model and 2) the type of HSUV data required by NICE (section 2.2, Box 1).

## 2.1. HEALTH STATES REQUIRED FOR COST-EFFECTIVENESS MODEL

To begin with, determine the distinct health states for a disease or condition pathway within the decision model that each requires HSUV data. It is likely that HSUVs for more than one health state will be required. In reality, this may be an iterative process, with relevant health states emerging as the model develops, and so the need for HSUV data may increase or change. For example, a new side effect or adverse event might be identified that requires an appropriate HSUV value. Figures 2 and 3 outline some of the health states considered in our case studies that may each require a HSUV within a decision model of osteoporosis or early breast cancer respectively. The case studies demonstrate that there are multiple health states

to take into account in these disease areas and thus multiple HSUVs may be required in a decision model.

An additional consideration is the time horizon of the decision model. Decision models consider what happens to patients over a period of time and so a series of HSUVs are required to reflect the changes in patients' health states in that time. For example in figure 2, we can see that HSUVs for health states relating to fractures may be required at different time intervals since HSUVs immediately post-fracture are likely to be different to those at one year post-fracture.

**Figure 2: Identified HSUV data needs for a decision model for osteoporosis treatment and prevention**

| | |
|---|---|
| Health states | Disease stage<br>• Pre-fracture (age and sex-matched norms)<br>• Established osteoporosis → With vertebral deformity / Without vertebral<br>Fracture<br>• With a history of fracture<br>• Without history of fracture<br>Fracture type<br>• Vertebral fracture (with clinical input)<br>• Hip<br>• Shoulder<br>• Wrist<br>• Multiple fractures<br>↕ Time post-facture<br>Non-osteoporosis health states<br>• Breast cancer[13]<br>• Atrial fibrillation[14]<br>• Bone loss in periodontal disease[15] |
| Population subgroups | Age group<br>Menopausal state<br>• Pre-menopausal<br>• Post-menopausal |
| Other considerations | Setting<br>• Nursing home<br>• Independent living |

**Figure 3: Identified HSUV data needs for a decision model for early breast cancer**

| | |
|---|---|
| Health states | • Disease state<br>o Response to treatment<br>o Stable<br>o Progression<br>o Terminal<br>o Not specified<br><br>• Time since diagnosis<br>o Under 1 year<br>o 1-5 years<br>o More than 5 years<br>o Time not mentioned<br>↕ Time horizon<br><br>• Treatment type<br>• Chemotherapy<br>• Hormonal<br>• Radiotherapy<br>• Treatment not specified<br><br>• Side effects<br>o Peripheral neuropathy<br>o Oedema<br>o Febrile neutropenia<br>o Sepsis<br>o Hypocalcaemia<br><br>• Recurrence<br>o No recurrence<br>o Recurrence<br>o Recurrence not specified<br><br>• Risk of recurrence<br>o No recurrence risk mentioned<br>o Risk <15%<br>o Risk ≥15% |
| Population subgroups | • Age (deviation from the mean of 46) |

### *2.1.1. Population subgroups*

Once the individual health states have been defined, consider whether there are any population subgroups that have sufficiently different characteristics that require subgroup-level HSUV data. Clinical subgroups might include different stage of disease, presence of co-morbidities, age-group or ethnic group. For example, in Figure 1, clinically relevant subgroups for health states in an osteoporosis decision model might be pre-menopausal or post-menopausal women. There may also be other considerations that need to be taken into

account for health states, such as population setting. For example, separate HSUVs may be required for adults with osteoporosis living independently and those living in a nursing home.

### *2.1.2. Intervention effects on HUSVs*

This TSD focuses on instances where you need HSUVs that relate to a condition of health state broadly and external to a particular trial. However there may be instances where you are particularly interested in the specific effects of an intervention or treatment on QoL, for e.g. a [particular adverse event profile or social impact of treatment e.g. diabetic treatment. In these instances, it might be more appropriate to use the values from the trial itself in the model. An interesting issue, which might be more for further investigation is recognizing that HTAs might both want to establish the values of utilities AND to measure the impact of interventions on utility values. The latter is where reviews of interventional studies might have a place, as RCTs do sometimes use EQ5D as an outcome measure.

### *2.1.3. Health states beyond primary condition*

Data may also be required for health states that extend beyond the health states defined in the 'P' of PICO of the clinical effectiveness review in the health technology assessment. For example, interventions given for one disease may positively or negatively impact on another disease. For example bisphosphonates for the treatment of osteoporosis may reduce the risk of breast cancer [13] and alveolar bone loss in periodontal disease[15]; and thus the health states 'breast cancer' and 'bone loss in periodontal disease' must be included in the decision model which necessitates HSUV data. However, there is also evidence that bisphosphonates can have a negative effect on the risk of atrial fibrillation[14], and thus this health state might also have to be taken into account. Similarly, where evidence cannot be found for a specific health state it might be useful to extend the scope to consider other similar conditions. For example if HSUVs could not be found for diabetic retinopathy, consider other eye conditions that impact on visual functioning in a similar way.

## 2.2. TYPE OF HSUVS REQUIRED

The requirements for measuring and valuing health effects in Section 5.4 of the NICE methods guide[4] (p.38-9) are presented in Box 1.

**Box 1: NICE Methods guide requirements**

- 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
- Data collected from condition-specific, preference-based measures should be presented separately
- The use of utility estimates from published literature must be supported by evidence that demonstrates that they have been identified and selected systematically.

The EQ-5D is the preferred method for the measurement and valuation of HRQL in adults. Health state descriptions must be provided from the relevant population (i.e. patients) who complete the EQ-5D, and using the UK TTO value set. Usually, the scope must be kept broad in relation to the type of HSUV data and the advantages to this approach are discussed in section 2.2.2. Sometimes, the scope may be narrowed to focus on EQ-5D if relevant and plentiful EQ-5D is known to be available.

### *2.2.1. Relevant EQ-5D data is available*

Firstly, assess the extent to which relevant EQ-5D data exists for the each of the health states in the decision model. If it can be confidently determined that relevant and plentiful HSUVs from the EQ-5D exist, then the scope may be narrowed to include EQ-5D data only. However, the review process would still entail a systematic identification and selection of evidence as outlined in section 3.

There are two key ways to quickly assess the extent of the EQ-5D evidence in a topic area: 1) Undertaking a scoping search and 2) Examination of previous HTA submissions to NICE or other health care decision agencies (such reports are usually made available online).

#### 2.2.1.1. Scoping search

A brief scoping search can be undertaken to locate published reviews. This can be limited to a few key electronic databases and searching using terms for each health state combined with terms for the EQ-5D. The idea here is to get an overview of the data and key reviews will be picked up if they are available. (More details on search terms and sources for searching for HSUVs are presented in section 3). The usefulness of a scoping search will depend on how straightforward it is to tell what HRQL instrument or approach is used to derive HSUVs within the study and whether a study is relevant to the decision model-neither of which may be apparent in the study abstract. The additional time required to undertake a scoping search and review the studies it retrieves needs to be balanced against the extra time and resources required.

Our case studies in osteoporosis and breast cancer found a number of HSUVs from the EQ-5D. Thus, where HSUVs are required for parameters of decision models that involve health states examined in the osteoporosis and breast cancer reviews (e.g. hip fracture, early breast cancer), it may be reasonable to limit the review scope to EQ-5D data only. However, the EQ-5D data must also be judged as ***relevant*** to the decision model.

### *2.2.2. Relevance and quantity of EQ-5D data unknown*

Usually the scope must be kept broad in relation to the type of HSUVs required. Typically the extent of EQ-5D HSUVs and their relevance to the health states in the decision model won't be known at the scoping stage; particularly since determining that a study contains HUSVs in a relevant population can often only be decided by examining the full-text article (see section 3.3 for further discussion), a task which goes beyond the brief scoping search function.

Narrowing the scope too early by focusing on one specific health HSUV instrument may result in no studies being selected for the HSUV review, and subsequent repeating of the review process for other HRQL instruments. A broad approach allows data from studies using non-preference based HRQL measures to be identified and mapped to the EQ-5D (TSD10[16]) or even provide other data where the EQ-5D is deemed inappropriate (TSD11[17]). In addition, keeping the review broad allows relevant contextual information on HRQL to be

identified as it is useful to see such EQ-5D data in the context of other important HSUV data (e.g. values derived from other instruments such as the HUI-3) or in the context of other more general HRQL data.

Following identification of a potential body of evidence for the review, decisions can be made on further refining or narrowing the scope with respect to the particular type of HSUV data. Section 3.3 will discuss the scope refinement in relation to the evidence base available.

#### *2.2.3. EQ-5D is not appropriate for the condition or effects of treatment*

The Methods Guide acknowledges that the EQ-5D may not be an appropriate measure to calculate HSUVs in all circumstances;[4] and stipulates that in such an instance empirical evidence be provided on why the properties of the EQ-5D are not suitable for a particular patient population or condition (e.g. properties such as validity, responsiveness and reliability). The issue of appropriateness of the EQ-5D is explored in more detail in another TSD within this series (TSD11[17]). There might also be values generated by non-preference-based measures which might be mapped onto the EQ-5D, (TSD10[16]) and also values from condition specific preference-based measures that may provide evidence for populating the models (TSD11[17]). The judgment about the usefulness of such data for the decision problem will be appraised by NICE.

## 3. IDENTIFYING AND SELECTING THE EVIDENCE

The Methods guide stipulates that where HSUV estimates are used from the published literature, there must be evidence that demonstrates they have been identified and selected systematically.[4] There is a lack of empirical evidence on the optimal approach for searching for HSUV data, with the methods guidance for reviews of clinical effects providing no guidance on searching for HSUVs.[6,7] There is also no known validated methodological search filter for HSUV data, unlike reviews of clinical effects where several search filters exist for types of study design, such as RCTs.

### 3.1. COLLECTIONS OF HSUVS

Published HSUV reviews (such as the osteoporosis review[8] or breast cancer review[9]) or reviews included as part of HTA submissions to NICE or other agencies can provide HSUVs. Locating an HSUV review can be an efficient way of identifying HSUVs for HTA

submissions to NICE. However, where a previous systematic review is used it is essential to establish it as of excellent quality, up-to-date and relevant to the scope. This requires some quality assessment and guidance on assessing the quality of HUSV reviews is provided in section 4. Most importantly, you need to be satisfied that the review has used systematic and transparent methods of identifying and selecting HSUVs. If so, the review can be updated by repeating the search process outlined in the review as required. It may be that only selected information is needed from a previous review, and so it is more likely that a review will be used as a source of studies rather than to provide an overall HSUV for a health state. Where HSUVs are taken from a previous review, some assessment of the relevance of the studies the values are taken from must be undertaken in relation to the current review question and to the NICE reference case (see section 4).

## 3.2. LITERATURE SEARCHING

Literature searching in systematic reviews of clinical effects is extensive and involves multiple methods for study identification. For HSUV reviews, we recommend using a sensitive search approach for identifying HSUV studies by development of an extensive list of search terms replicated on a number of electronic databases, supplemented by other methods of study identification such as reference list checking. There is an added problem that the health states that require HSUVs may increase or change, thus requiring further searching. Similarly, if HSUV data cannot be identified for a health state it might be necessary to undertake further searching for evidence in a related condition or for HRQL data from condition-specific measures. Therefore, searching for HSUV review is likely to involve undertaking a sequence of searches and further searching as health states emerge. Therefore keeping careful records of each search undertaken (and selections from the search results) makes the search process transparent.

There are three particular issues to consider when searching for HSUVs: a) search terms to use, b) sensitive versus precise searching and c) where to search. In section 3.2.4 we present the search strategy used in the osteoporosis case study.

### *3.2.1. Search terms*

The terminology to use for the health states aspect of the literature search should be fairly straightforward to identify. However, it's important to cover all the health states required in the decision model. To a certain extent this may be iterative as new relevant health states are identified and require HSUVs, new searches will be undertaken. In our case study examples, there were a number of relevant health states including 'metastatic breast cancer', 'adverse events' and 'hip fracture' etc.

Terminology for HSUV terms is more problematic with two issues: 1) relevant subject headings and 2) relevant words appearing in titles and abstracts of records.

#### 3.2.1.1. Subject headings

The thesauri in Medline (MeSH) and Embase (EMTREE) provide little coverage of this topic with no available dedicated thesauri terms for common HSUVs (EQ-5D, SF-6D etc). However some terms seem to be consistently applied and HSUV studies are typically indexed under broader concepts such as **Quality of Life** or **quality adjusted life year**; terms which are not directly relevant to HSUVs. Paisley et al[1] identified a cross-sectional sample of 300 records from Medline retrieved using HSUV related-free-text terms. The indexing terms in those records were examined to determine the most frequently applied MeSH heading[1] and these are listed in descending order of frequency in Box 2.

We reviewed the subject headings attached to the studies included in the osteoporosis case study review. Of the 28 included studies, 24 were indexed on Medline (irrespective of their method of identification in the original review). Of the 24 studies, all but one was assigned the MeSH term '**Quality of life'**. The next most frequently assigned MeSH term was '**Questionnaires**' being assigned to eight of the 24 studies. However, most of the other MeSH terms listed in Box 2 were not assigned to the studies included in the osteoporosis case study, or were assigned to two or three studies at most.

Whilst searching using '**Quality of life'** or '**Questionnaires**' as MeSH terms might appear to be a useful method of identifying relevant studies, there is a trade off to be made between sensitivity and precision. Both these MeSH terms are generic and whilst their use in an electronic database search will maximise sensitivity of the search and reduce the risk of missing relevant records, it will also increase the size of the result set when combined with

some disease areas, within which lies a small number of relevant studies. (i.e. lower precision).

**Box 2: Frequently applied MEDLINE MeSH terms to HSUV studies (based on examination of 300 records) [1]**

Quality of life
Questionnaires
Psychology (subheading)*
Health status
Health status indicators
Activities of daily living
Health surveys
Quality adjusted life years
Treatment outcome
Psychometrics

*Subheadings are terms that can be added to subject headings to refine their meaning

3.2.1.2.Free-text searching

Relevant free-text terms fall into three categories. Firstly there are general terms (e.g. QALY, HSUV); secondly instrument-specific terms (e.g. EQ-5D, SFD-6D etc); and lastly terms that relate to the associated methods of elicitation (e.g. standard gamble, time trade off). When using free text terms, it is important to take into account that terms may be referred to or spelled in different ways. This is particularly pertinent for search terms for HRQL instruments, which may be referred to by their full name or abbreviated name (e.g. EQ-5D, euroqol, euro qol, eq5d etc). Figure 4 lists some free-text terms for use in electronic database searching for HSUVs showing the importance of synonyms, abbreviations and spelling variants. However, searching using free-text terms relies on terms relating to HSUVs being present within the title or abstract of studies. Since HSUV data are often reported as secondary or tertiary outcomes, HSUV-related terms (e.g. EQ-5D, HSUVs) may not be mentioned in the abstracts, thus these studies will not be retrieved by searching in this way.

We reviewed the free-text terms attached to the studies included in the osteoporosis case study review. Seventeen of the 24 studies indexed on Medline included the words '**Quality of life**' in the title; five of the remaining seven contained '**Quality of life**' in their abstracts. All

but two of the 24 studies included free-text terms relating to the specific quality of life instrument used to generate the HSUVs in their titles or abstract.

**Figure 4: Common free-text terms for electronic database searching for HSUVs**

quality adjusted life

quality-adjust-life (note not all databases can hope with hyphens)

(qaly$ or qald$ or qale$ or qtime$)

disability adjusted life

daly$

(sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six)

(sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six)

(sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve)

(sf6D or sf 6D or short form 6D or shortform 6D or sf six D or sfsixD or shortform six D or short form six D)

(sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty)

(euroqol or euro qol or eq5d or eq 5d)

(hql or hqol or h qol or hrqol or hr qol)

(hye or hyes)

health$ year$ equivalent$

utilit$

(hui or hui1 or hui2 or hui3)

disutili$

rosser

quality adj2 wellbeing

qwb

standard gamble$

SG

time trade off

time tradeoff

tto

**Key**

$ =truncation (In some databases this is *) e.g. utilit$ searches for utility or utilities

adj= adjacency operator. e.g. adj2= within two words of each other

As in reviews of clinical effectiveness, the main method of identifying studies containing HSUV data is searching of electronic databases. As well as searching health-related databases such as MEDLINE, EMBASE and CENTRAL; and conference proceedings (e.g. ISI

Proceedings) there are several specialist health economics resources that may be useful to search. These include:

- Cost-effectiveness Analysis Registry (formerly known as the Harvard CUA database)[18]
- Centre for Reviews and Dissemination (CRD) databases: NHS EED and HTA[19,20]
- Health Economics and Evaluation database (HEED)[21]
- The EQ-5D website[22] (and other instrument sites)
- The MAPI Institute website[23]
- Submissions to NICE or other health care decision agencies
- Research Papers in Economics (RePEc)[24]
- Patient-reported outcome and quality of life instruments database (PROQOLID)[25]

#### 3.2.2.2. Supplementary search techniques

Searching electronic databases is just one method of identifying the evidence for a systematic review. Reviews of clinical effects routinely identify studies by other methods including reference list checking and contact with experts. Literature searches to identify published literature on HSUVs should ideally include other methods of identifying studies, particularly given their usefulness in locating relevant studies (see section 3.2.3).

Both case studies demonstrate the importance of supplementary search techniques in the identification of studies in HSUV reviews. The osteoporosis review gives a full account of where studies were identified, with 13 of the 28 studies identified through the electronic search; accounting for less than 50%. A further eight papers were found by scrutinising the reference lists of included studies, four by contact with experts and two from a previous systematic review. The osteoporosis case study demonstrates how valuable supplementary methods are. The extent to which there are low levels of sensitivity in electronic searches in the identification of HSUVs within other areas of health care is not known.

Other methods of identifying studies include citation searching and author searching. Citation searching involves taking key relevant papers and identifying subsequent studies that have since cited the paper(s). Where it is evident that a key author exists in a field, it may be useful to search for further relevant publications by the author.

### *3.2.3. Sensitive or precise?*

For reviews of clinical effects, a sensitive approach to searching is standard; that is an extensive search strategy is devised which provides a range of synonyms for every concept to be included in the search. The emphasis is on maximising sensitivity of electronic database searching so reducing the risk of missing relevant records, whilst increasing the size of the result set. In addition, searches for clinical effectiveness reviews are often limited by using a validated RCT filter.

We recommend using a sensitive approach to electronic database searching for HSUV reviews, however there are there are drawbacks in adopting a sensitive approach. This is in part due to fewer resources and less time available to cope with large results, which is likely to be of low precision given the problems in indexing HSUV studies. However, the aforementioned problems with non-specific thesauri terms and the reliance on suitable free-text terms appearing in titles or abstracts mean that adopting a sensitive approach to searching does not necessarily mean that all relevant records will be found. In fact, our case studies demonstrate that not only were a large proportion of references for the two HSUVs reviews located by means other than electronic database searching, this was despite the fact that often studies identified by other means (e.g. by reference list checking) were in fact indexed on MEDLINE and thus were there to be identified by the electronic database search.

Adopting a precise or focused approach to searching has its own associated problems. By narrowing the electronic search too early, for example by using search terms to locate only those studies on the EQ-5D, may mean no relevant studies are identified if such data does not exist or are not relevant to the review. Furthermore, it will usually be difficult to determine at the literature search stage what data are available. It may even be difficult to determine which HRQL instrument has been used without seeing the full-text of the paper because of unclear or unspecific reporting in the abstract.

### *3.2.4. Example of a search strategy to identify HSUVs-osteoporosis review*

Both of the case studies described in this TSD adopted a sensitive approach to electronic database searching i.e. with extensive search strategies and large result sets. Appendix A.1 presents the search terms used in MEDLINE for the osteoporosis review. Below we examine

the search strategy, including the search terms used and the methods/sources searched, highlighting the strengths and weaknesses of the approach.

3.2.4.1. Search terms

**Health state terms**

- Health states: male and female adults with conditions related to osteoporosis: established osteoporosis, vertebral describe osteoporosis
- Steps 1-4 are terms to describe osteoporosis
- Steps 5-14 are terms to describe bone density
- Steps 15 to 46 combine terms describing fractures with terms describing hormone replacement therapy or the menopause.
- Step 46 combine the terms for the health states with OR.

**HUSV terms**

- HSUVs: empirically estimated using a recognised valuation technique for obtaining empirical HSUVs (typically VAS, SG or TTO).
- Include broad subject heading terms such as **Quality of life** (steps 48 and 49) or broad free-text terms such as **quality of well-being** (step 61 and 62)
- Include specific instrument terms: SF-36 (step 52), EQ-5D (step 54), HUI 1-3 (step 60) and SF-6D (step 71)
- Include terms relating to method of utility elicitation (steps 65-70)
- Other general terms used such health utilit$, health state$ preference$ and health state$ utilit$ ($-denoted truncation, see critique below for explanation).

**Limits**

- Steps 76-79 included terms to exclude records that were letters or editorials
- Publication date limited to post-2000 since a previous review had already been undertaken (step 82)

**Sources**

- Ten electronic databases searched (including searching for conference proceedings): Cochrane controlled trials register (Central), Cochrane database of systematic reviews, EMBASE, Science Citation Index, MEDLINE, MEDLINE in process, NHS

Database of Abstracts of Reviews of Effectiveness (DARE), NHS Economic evaluation database (NHS EED), Index to Thesis, ISI proceedings

- Hand-searching of two journals: Osteoporosis International and Journal of Bone and Mineral Research
- Reference list of relevant studies checked.
- Contact with experts with a clinical and economic background in osteoporosis research.

Table 2 records the methods by which studies were located for the osteoporosis reviews, emphasising the need for both a sensitive search approach including using an extensive list of search terms in an electronic database searching, reference list checking and other methods and contact experts.

**Table 2: Studies included in the osteoporosis review**

| Method of identification | Number |
|---|---|
| Electronic database searches | 13 |
| Reference lists | 8 |
| Contact with experts | 4 |
| Previous reviews | 3 |

## Critique of search approach

The search terms used were extensive and resulted in a very comprehensive search strategy used to search the electronic databases. A good mix of subject headings and free-text terms were utilised. Where free-text terms were used, truncation was used (denoted by the $ symbol), for e.g. utilit$ would find utility or utilities. This search would be good at picking up HSUVs from specific measures such as the SF-36, EQ-5D, HUI 1-3 and SF-6D. The broad subject heading **Quality of life** was used as a search term and thus this increased the likelihood of picking up more 'general articles' on quality of life in osteoporosis and related conditions. The advantage of using this **Quality of life** meant that HRQL data from condition-specific HRQL measures might be identified (although we would recommend undertaking a separate search using condition-specific HRQL measure terms if these were

required). However, using the term **Quality of life** increases the likelihood of retrieving non-relevant articles and a large result set of lower precision. There were some redundant terms, for e.g. health utilit$ (step 59) and health state utilit$ (step 72) might have been searched under one step: utilit$. In addition some steps could have been further refined by including adjacency operators (e.g. health adj2 state adj2 value$.tw. for step 73).

A number of electronic databases were searched, although it may have been helpful to search specialist health economics resources. Several strategies were used to identify studies and the importance of this is highlighted in table 2, where over half the included studies came from methods other than electronic database searching (reference list checking, contact with experts and previous reviews). The limits applied were justified and served to reduce the result set to a more manageable number.

### 3.3. SELECTING STUDIES AND INCLUSION/EXCLUSION CRITERIA

Selecting studies for inclusion in HSUV reviews differs from the process used in reviews of clinical effects due to the need to further refine the scope and inclusion/exclusion criteria of the type of HUSV data required as the evidence base emerges. As in reviews of clinical effects, titles, abstracts and full texts are screened against pre-defined inclusion and exclusion criteria. At each stage, studies are rejected if they do not meet one or more of the eligibility criteria.

However, the process of making inclusion and exclusion decisions in HSUV reviews can be difficult at title and abstract level due to the problems with the standard of reporting within study abstracts. There is an increased risk of failing to select a relevant study and/or increase the number of studies that require full-text screening. We recommend undertaking some brief data extraction of three key details in order to help the decision-making process and assess the nature of the evidence based:1) Population that is the subjects of the health state, 2) Details of the approach used to describe the health state and 3) Valuation methods

Indeed, the osteoporosis review noted that 13 studies were identified by reference list checking when in fact eight of these studies were indexed on Medline with terms that were used in the review's electronic search strategy. Thus, these studies appear to have been missed during the study selection stage.

### *3.3.1. Preliminary data-extraction*

Once studies have been identified from the literature search, as potential includes, a preliminary sift can be undertaken to remove any obviously non-relevant records and focus on the studies that might contain HSUVs for the required health states in the decision model. From this initial sift, it will be possible to get a feel for the quantity and nature of data and thus you can start to refine the scope by weighing up the characteristics of the possible studies in order to make a choice of the studies to be included. To do this, it is useful to undertake some preliminary data extraction to extract the following three key details:

- Population or the subjects of the health state (e.g. age, sex, disease severity): this should reflect those in the model and are the individuals who describe the health state
- Details of the approach used to describe the health state- a) Vignettes or scenarios, b) Generic multi-attribute health state descriptive systems e.g. EQ-5D, SF-6D or c) Direct measurement by TTO, SG, VAS (Note if proxy values are used).
- Valuation methods: Who? (E.g. general public) and How? (HSUV elicitation technique such as TTO, SG, VAS)

Based on the findings from this preliminary data extraction stage, a picture can be built up about the nature and quantity of evidence available, which in turn helps to develop inclusion and exclusion criteria. Think of this as an exercise to describe the evidence available in order to make decisions about where to refine the inclusion and exclusion criteria.

For NICE HTA submissions, it is useful at this stage to consider the following:

- Are studies with HSUVs derived from the EQ-5D data available and how many?
- Are EQ-5D studies relevant to the health states in the decision model?

#### 3.3.1.1. Relevant and plentiful EQ-5D data

When relevant EQ-5D data is available and appropriate to use, the inclusion criteria can be narrowed to include only EQ-5D evidence. Studies that are relevant in terms of health states to the decision model and containing HSUVs derived from other measures can be excluded, although keep a record of such studies and the reason for exclusion (i.e. to be used only for contextual information and not for values in decision model).

3.3.1.2. No relevant EQ-5D data available

Where EQ-5D data is appropriate but relevant data are not available, seek to identify data from other measures such as the SF-36 and condition-specific measures which can be mapped onto the EQ-5D, again documenting reasons for inclusions/exclusions as appropriate (TSD10[16]).

3.3.1.3. EQ-5D inappropriate

Where EQ-5D is inappropriate or where there are no data that can be used to generate EQ-5D using published mapping functions then the next step is to look amongst the records to identify for studies that provide HSUVs from 1) Generic preference-based HRQL measures (e.g. HUI3), 2) Condition-specific HRQL measures or 3) Vignettes. The use of alternative sources of data is considered in TSD 11.[17]

### *3.3.2. Rationale for preliminary data extraction*

Whilst preliminary data extraction, can look like an extra stage in process, it performs two important functions. Firstly, this stage will ultimately save the reviewer time during the review process by reducing the number of studies that require full data extraction (see section 5). Secondly, this process provides a record of why studies have been included or excluded. From a review of clinical effects viewpoint, failing to provide pre-defined inclusion and exclusion criteria goes against one of the core principles of a systematic review i.e. the reduction of selection bias. However, we are not suggesting that this process is a license to include any HSUVs based on the choice of the reviewer. Inclusion and exclusion are specified at the scope stage; this process is a way of refining the inclusion criteria further in relation to the type of HSUVs according to the evidence base.

There must be full and explicit justification for inclusion of any study that provides a HUSV for the decision model and for exclusions of studies that appear to be suitable for inclusion. For example, it is reasonable to exclude studies that derive HSUVs from health states described by the vignette approach where plentiful data exists from patient reported outcome measures (such as other generic or condition specific preference-based measures) (See TSD 11[17] on the use of these alternatives). Similarly, a study might be included because its population is of high relevance to the population/health state within the decision model. However, there must be sound and explicit reasons for recorded and presented these choices.

### *3.3.3. The case study inclusion and exclusion criteria*

The inclusion criteria were kept broad in both the case studies we present, resulting in reviews that reported on HSUVs from a wide range of approaches and a number of HRQL instruments. The osteoporosis case study specified its population but not the individual health states prior to scoping, searching and selecting the evidence for the review e.g. wrist fracture; the breast cancer review included a broad population. Neither case study pre-specified the individual instruments used to describe the health state profiles (e.g. EQ-5D). However, the reviews both specified that studies must estimate HSUVs using a recognised valuation method (typically VAS, TTO or SG). Such an approach might be too broad for use in a decision model, which as the nature of the evidence becomes apparent may narrow the inclusion criteria according to instruments used to describe the health states and derive HUSVs and health states in the decision model.

**Table 3: Case studies' inclusion criteria**

| **Osteoporosis review**[8] | **Breast cancer review**[9] |
|---|---|
| • Adults > 17 years of age<br>• Men and postmenopausal women suffering primary or secondary osteoporosis<br>• Empirically estimated HSUV using a recognised valuation technique- SG, TTO or VAS<br>• English language or translation. | • Adult population with breast cancer<br>• One original, unique HSUV value, derived via SG, TTO or VAS<br>• Details of elicitation technique and respondents described<br>• English language or translation.<br><br>Nb: Studies therefore excluded HSUVs based on judgement, either of a non-specified clinical staff or of the author. |

Finally, as discussed in Section 2, there may be several sets of HSUV data corresponding to different HSUV data required for the decision model, and so the output at this stage may be several sets of references. For example in the breast cancer review, studies were ultimately split into six categories: screening-related states, preventative states, adverse events in breast cancer and its treatment, non-specific breast cancer, metastatic breast cancer and early breast cancer.

## 3.4. DOCUMENTING THE REVIEW

### *3.4.1. Searches*

It is essential to record details of the search strategy so that the method of study identification is transparent and reproducible. In the NICE manufacturer/sponsor HTA submissions template[26] section 6.4.5 request an account of the search strategy including rationale for terms used in the search strategy and any conclusion and exclusion criteria. (See Appendix A.4 for NICE template headings). NICE also request that the search strategy used is provided in section 9.12, appendix 12 of the template. Details of the sources searched would also be useful here.

However as new health states requiring HSUVs emerge as the decision model develops, it might be necessary to undertake further searching in addition to the 'main search' which each require transparent documentation (including the selections made from each search). These could be documented in a similar way as above, or where searches are precise and involve reduced number of search terms (for e.g. a search using terms for a health state and a condition specific measure), they might be usefully reported in a table similar to Table 4 below:

**Table 4: Table to record searches undertaken for new health states**

| # | Search terms | Source | Hits | Outcome of search |
|---|---|---|---|---|
| e.g. | ((Schizophrenia adj2 Quality adj2 Life adj2 Scale) or SQLS).tw. | PsycInfo | 136 | Number included at title sift= 23<br>Number included at abstract sift= 3 |
| .. | .. | .. | .. | .. |

### *3.4.2. Inclusions and Exclusions*

A flow diagram can be useful to present the decisions made during each stage of the selection process such as total number of search results, records included/excluded at title level, record included/excluded at abstract level and so on (see below for the flow diagram for the osteoporosis case study).

However, to capture the decision making process, an audit trial of the decisions to include and exclude studies for the review is very useful and helps to justify the selection of HSUVs in the decision model. For this, it is only necessary to record reasons for those studies that might be considered as possible include: e.g. where HSUVs are reported for the exact or similar population to that of the health states in the decision model. The rationale behind this is to be explicit for reasons for exclusion for those studies that at first glance a reader might question the decision for exclusion. Although, there is no natural position for this information in the NICE manufacturer's submission template, it might usefully be included in section 6.4.9 with the table detailed below:

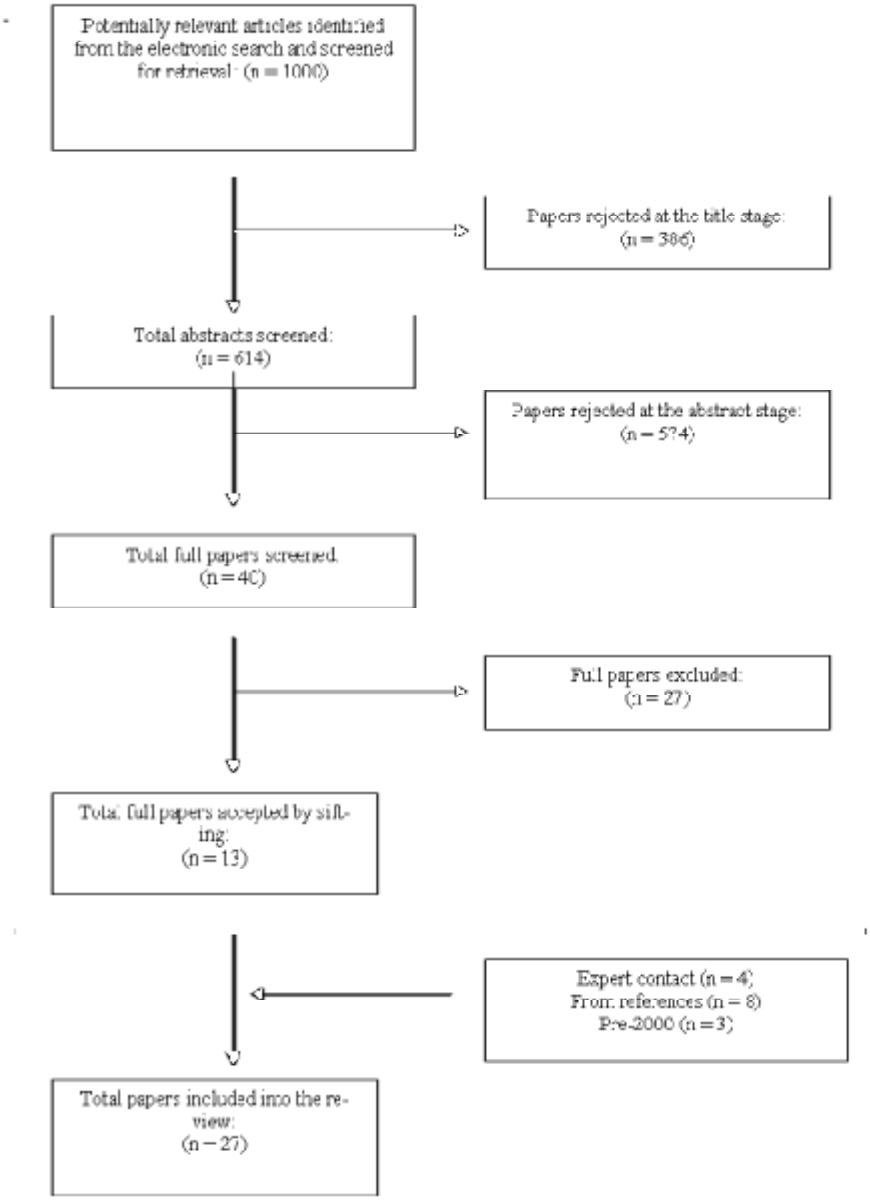
**Table 5: Table to record exclusion decisions**

| Study | Stage excluded at | Reason for exclusion |
|---|---|---|
| Bloggs, J (2009) | Prelim-data extraction | Contains SF-36 data, plenty of EQ-5D data is available |
| James, C (1998) | Full data extraction | Update of this study is available |
| Smith, A (2003) | Quality and relevance assessment | Population is 20 years younger than that included ion the model. |
| ... | ... | ... |

Included studies can be presented as per the table in the NICE manufacturer/sponsor submission template within Table B15 (section 6.4.9)

The osteoporosis review gives a clear and explicit account of where studies have come from by providing a flow diagram (see figure 5)

**Figure 5: Figure 1 Statement flow diagram of study selection and exclusion[8]**



# 4. QUALITY AND RELEVANCE ASSESSMENT

## 4.1. QUALITY ASSESSMENT OF HSUV REVIEWS

Where reviews of HSUVs (published or previous HTA submissions) are used as a source of studies, it must be established that the review is of excellent quality, particularly in relation to the methods used to identify and select HSUV studies which must be systematic and justified. The **Critical Appraisal Skills Programme** (CASP) systematic reviews checklist[27] has been adapted (the CASP reviews checklist is provide in Appendix A.2) for use in the quality assessment of reviews of HSUVs.

**Figure 6: Adapted CASP quality assessment checklist[iv] for HSUV reviews**

For the following questions, answer Yes, No, Partly or Can't tell:

**1. Did the review ask a clearly-focused question?**

*Consider if the question is focused in terms of:*

- *Population describing the health states (ideally patients)*
- *Population valuing the change in HRQL (ideally public)*
- *Method of elicitation (ideally choice-based method e.g. TTO)*

**2. Did the review include the right type of study?**

*Consider if the included studies:*

- *address the review's question*
- *are appropriate studies*

**3. Did the reviewers try to identify all relevant studies?**

*Consider as a minimum:*

- *Were a number of electronic databases searched? (ideally clinical and specific health economic )*
- *Were reference lists scrutinised for retrieved references?*

[iv]The CASP Systematic Reviews checklist has been reproduced in Appendix x and adapted for use in HSUV reviews (in section 4 of the TSD) with the kind permission of CASP at Solutions for Public Health. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the Public Health Resource Unit. © Public Health Resource Unit, England 2006

*Ideally, but not mandatory, consider that the search methods should involve:*

- *personal contact with experts*
- *search for unpublished studies*
- *citation and author searching*

**4. Did the reviewers assess the quality of the included studies?**

*Consider the:*

- *Sample size*
- *Respondent selection and recruitment*
- *Inclusion/exclusion criteria*
- *Response rates to instrument used to*
- *Numbers (%) lost to follow-up*
- *Are reasons provided for any loss to follow-up?*
- *How is missing data from the instruments used to describe the health states dealt with? Is the method rigorous?*
- *Any other problems with the study*

**5. Did the reviewers assess the relevance of the included studies to the review question?**

- *Population describing the health states (ideally patients)*
- *Population valuing the HRQL (ideally public)*
- *Method of elicitation (ideally choice-based method e.g. TTO)*

**6. If the results of the studies have been combined, was it reasonable to do so?**

- *the results of each study are clearly displayed*
- *the results were similar from study to study (look for tests of heterogeneity)*
- *the reasons for any variations in results are discussed*

**7. How are the results presented and what is the main result?**

- *Is there a full account of why studies were excluded? (includes factors relating to relevance)*
- *Is there are full justification of why studies were included?*
- *how the results are expressed (descriptive statistics or coefficients of a model.)*

**8. How precise are these results?**

*Consider:*

- *if a confidence interval were reported. Would your decision about whether or not to use this* intervention *be the same at the upper confidence limit as at the lower confidence limit?*
- *if a p-value is reported where confidence intervals are unavailable*

**9. Can the HUSVs be used in the health states in your decision model?**

*Consider*

- *How relevant the population describing the health state is to the health state sin the decision model*
- *Have all subgroups been considered e.g. age, disease severity, setting*
- *Do the HSUVs match the NICE reference case?*
- *How do the results need to be modified for the decision model?*

## 4.2. QUALITY ASSESSMENT OF INDIVIDUAL HSUV STUDIES

Reviews of clinical effects typically involve an assessment of the quality of included studies, typically in terms of assessing the risk of bias in relation to study design or conduct.[7] For HSUV reviews to inform decision models, there is a need to assess the relevance of the evidence as well as the quality. Therefore, for each study to be included in the review there needs to be a full quality and relevance assessment.

In HSUV reviews, quality of included studies can be difficult to assess as there are no agreed reporting standards for these types of studies. HSUV studies do not fall into a particular study design; simply choosing a quality assessment checklist based on the study design may not be appropriate. Where HSUV data are secondary or tertiary outcomes (which is often the case within trials), it is important to consider the possible uncertainty in the HUSV results since the study may not have been powered and designed according to non-primary outcomes.

Based on experience and also incorporating the approach used in the case studies, quality assessment of HSUV studies might usefully focus on respondent selection and recruitment, inclusion and exclusion criteria and a description of the background characteristics of the sample population from whom values are obtained. It is also important to examine the response rates of the measures used to derive the HSUVs and any loss to follow-up, particularly where values are collected over time for e.g. response rates at baseline compared

with follow-up. Czoski-Murray et al[28] found that in populations with hip fracture, those who had a full data set (i.e. HRQL values for all timepoints in the study) had a better recovery from hip fracture than the full study sample. Lastly, it is helpful to make a note of any further potential problems with a study and their potential impact on the validity and robustness of the HSUVs. Box 3 outlines the essential quality assessment criteria and what to consider for each.

**Box 3: Key criteria to consider in quality assessment of HSUV studies**

| Criteria | Consider… |
|---|---|
| Sample size | This is not an exclusion criteria, but the precision of the estimate should be reflected in the variance around any estimate used in a model |
| Respondent selection and recruitment | Does this result in a population comparable to that being modelled? |
| Inclusion/exclusion criteria | Do these exclude any individuals that might (e.g. the very elderly >80 years old are often not included in studies) |
| Response rates to instrument used to | Are response rates reported and if so, are the rates likely to be a threat to validity? |
| Loss to follow-up | How large is the loss to follow-up (e.g. one year after a fracture) and are these reasons given? What are these likely to threaten the validity the estimates |
| Missing data | What are the levels of missing data and how are they dealt with? Again, could this threaten the validity of the estimates |
| Any other problems with the study | Example: Relevance of location (e.g. if patients recruited in non-UK country) |

The relevance of the data to the decision model and agency to which the model will be submitted is as important as quality assessment. The relevance of the data to the decision model will involve comparing the participant characteristics in the individual HSUV study and the population being modelled. For example, our osteoporosis review identified that institutionalised adults were excluded from studies, and thus the HSUVs from this review could not be applied to this population in a decision model.

Assessing relevance according to the agency to which the model is being submitted requires detail on the HSUV data collection methods. Box 4 suggests some questions to ask when

assessing relevance of HSUV data, with the column on the right specifying the relevance criteria in relation to NICE.

**Box 4: Relevance assessment criteria for NICE reference case**

| Relevance questions | Requirement for NICE |
|---|---|
| Do the population characteristics (e.g. age, sex, co-morbidities, diagnosis, severity of disease) in the study match those modelled and those described in the decision problem of the review? | If the answer is no, there are techniques available that may be able to adjust the values to make them more relevant to the decision model (TSD12[29]) |
| What instrument is used to describe the health states? | Generic preference-based instrument, preferably EQ-5D |
| From which population is the change in HRQL undertaken? | Directly from the patient |
| From which population is the valuation of changes in patients' HRQL undertaken? | General population |
| What is the technique used to value the health states? | Choice-based method such as TTO |

# 5. DATA EXTRACTION

A more complete data extraction (than that outlined in section 3.3.1 (preliminary data extraction)) needs to be undertaken for those studies identified at the selection stage for inclusion in the review. The purpose of the 'full' data extraction stage is to i) help further inform the inclusion/exclusion of studies by weighing up the characteristics across candidate studies, ii) identify where it may be possible to synthesis HSUVs and identification of factors that need to be considered in interpretation of a synthesised HSUV (i.e. heterogeneity) and iii) identify the data that will inform how HSUVs may need to be modified for use in the health states of the decision model.

Data extraction for HSUV reviews, as for clinical effectiveness reviews, involves design of a data extraction form to pre-define exactly what data to extract and should ideally be piloted to ensure it is collecting the necessary data.[6,7] The CRD guidance outlines the types of information to be extracted for clinical effectiveness reviews,[6], and much is relevant to data extraction in HSUV reviews. For example, general information such as author or country of publication; study characteristics such as inclusion/exclusion criteria; participant characteristics such as age, sex, disease characteristics and study setting. Information relating

to outcomes differs because different types of data are collected and unique/specific techniques are involved in HSUV collection.

The process of data extraction in clinical effectiveness reviews is ideally undertaken by two independent reviewers.[6,7] However, more often data extraction is undertaken by one reviewer and checked by a second reviewer. A similar method might be applied to HSUV reviews depending on the time and resources available.

Based on experience and by examining the types of data extracted for the case studies, a sample data extraction form is suggested-see appendix A.3. Some modifications may be necessary, particularly in relation to disease-specific information.

# 6. SELECTION AND SYNTHESIS OF HSUVS

NICE requires that any selection process is transparent and justified, though it does not have to involve formal quantitative synthesis. However, in some situations a formal synthesis may help to justify the values selected. Methods for selecting and synthesising HSUVs for use in a decision model depends on the availability of data that are relevant to the population being considered in the decision model and the reference case requirements of NICE. There are three situations to consider when selecting.

## 6.1. ONE SET OF RELEVANT HSUVS

Where there are only one set of HSUVs shown to be relevant to the population being modelled and that meet the NICE reference, then these values should form the central estimates for the states used in the model. The identification and extraction processes will be sufficient to support the selection of the values used. However, it is advisable to still use values using other sources in order to examine the sensitivity of the ICERs to the HSUVs used in the model.

## 6.2. MULTIPLE SETS OF RELEVANT HSUVS

Where there are more than one set of HSUVs meeting the criteria of the review, then the selection of values used in the model needs additional justification. It might be possible to show that one set of values dominates other sets in terms of relevance to model and meeting the NICE reference case. More difficult will be situations where one set of values is better on some criteria but worse on others (e.g. EQ-5D values come from a more clinically relevant population, but non-UK). The importance of different criteria should be fully considered in a narrative review to support the final selection. Whatever choice is made, it is important that sensitivity analyses are undertaken and presented in order to better understand the impact of the choice.

Where more than one set of relevant values are sufficiently homogenous (i.e. collected from the same patient population using the same instrument and valued using the same UK value) then pooling should be considered as a way to improve the precision of the estimates of the mean HSUVs and their variances. The osteoporosis review of HSUVs reported in this guide found a number of studies reporting HSUV values using the EQ-5D in hip fractures.[8] An estimate of the overall QALY loss in the first year following a hip fracture was estimated by pooling EQ-5D values provided across five studies weighted using the inverse of the variance and by sample size (though these two methods provided very similar results). Pooling methods such as these are simple to undertake and may be acceptable provided the populations are sufficiently homogenous and the values were obtained using the same instrument (EQ-5D) and value set.

## 6.3. NO DIRECTLY RELEVANT VALUES

Where there are no directly relevant values, such as the reference case has not been used or the patient group is not appropriate for the model, values still need to be selected and justified. Some form of synthesis may help in this situation in order to better understand the causes of variation and hence provide support for the values selected. In the other case study,[9] there were 118 useable values from 19 studies in metastatic breast cancer (MBC). These were values obtained for a number of states that varied according to degree of disease progression

(response, stable, progression and terminal), treatment (e.g. chemotherapy, radiotherapy, and hormonal therapy) and side effects (e.g. peripheral or severe neuropathy with or without treatment interruption, oedema with or without treatment interruption, febrile neutropenia with or without hospitalisation), and age. They also varied in terms of methods of eliciting values, with only eight (7%) studies across different disease states reporting EQ-5D values. To make the most of these data meta regressions were performed for MBC and early breast cancer (EBC) using simple ordinary least squares (OLS) weighted by the standard deviation and the sample size of each HSUV estimate. The dependant variable was all HSUVs and the independent variables included the various sub-states and the methods of valuation. The analyses accounted for clustering within study. The sign of the coefficients associated with the various sub-states were usually as expected and the R-squared was generally quite good at 0.844 to 0.883. However, the methods of valuation were found to have a major impact on the results (such as whether patient or general population values were used) alongside those associated with the condition and its treatment. There may also have been socio-demographic and other clinical differences not controlled for in the models and aspects of methods not taken into account due to lack of observations. In contrast to the osteoporosis review,[8] the authors did not feel able to present a definitive set of health state values on the basis of these analyses. There was more success in the areas of HIV/aids and stroke where the authors used hierarchical linear models to estimate values across states for those conditions that can be helpful in cost effectiveness models.[30,31]

While the synthesis of clinical parameters is a well developed area of research, there has been little research into the synthesis of HSUVs. The problem until recently has been in obtaining any relevant HSUVs. With a growing literature of values there are increasingly situations where an analyst will have a number of values to choose from. NICE does not require a formal synthesis of HSUVs, but it may be helpful in maximising the reliability and precision of any estimates and providing a justification for the values selected.

## 6.4. MODIFICATIONS TO HSUVS

Finally there is a related issue of how to present the values for use in economic models. HSUVs often come from different patient groups to those in the model in terms of socio-demographic variables (e.g. age or sex), condition severity and the prevalence of co-morbidities. Furthermore, published HSUVs often come from cross-sectional data, when

what economic models require are estimates of how a condition and its treatment impact on HRQL (e.g. the lower scores on a patient with a hip fracture may not be due to the fracture, but also due to a lower pre-fracture score) change with treatment and its associated impact on patients' health (such as reductions in event rates). The analyst must adjust available data for these factors, and there are a range of methods for doing this that are examined in TSD12.[29]

# 7. PRESENTATION OF RESULTS

NICE provides a template for manufacturer/sponsor HTA submissions[26] which provides headings where most (but not all) of the relevant information from an HSUV review can be organised. Headings from the NICE template can be found in Appendix A.4

## 7.1. CHARACTERISTICS OF INCLUDED STUDIES

Providing detailed information on the characteristics of each study as well as the HSUV data is essential to allow readers to understand why values may differ across studies. In addition, providing such information allows readers to assess the relevance of the HSUVs to their own context, and in the case of NICE HTA submissions relevance to the NICE reference case. Such information is usually presented in section 6.4.6 (p.39-40) of the NICE template for the manufacturer's submission and as specified in the template includes the following elements:

- Population in which health effects were measured.
- Interventions and comparators
- Response rates.
- Adverse events
- Method of elicitation
- Mapping
- Consistency with reference case
- Appropriateness of the study for cost-effectiveness analysis
- Information on recruitment
- Sample size
- Description of health states
- Appropriateness of health states given condition and treatment pathway
- Method of valuation
- Uncertainty around values
- Results with confidence intervals

## 7.2. HSUVS USED IN THE DECISION MODEL

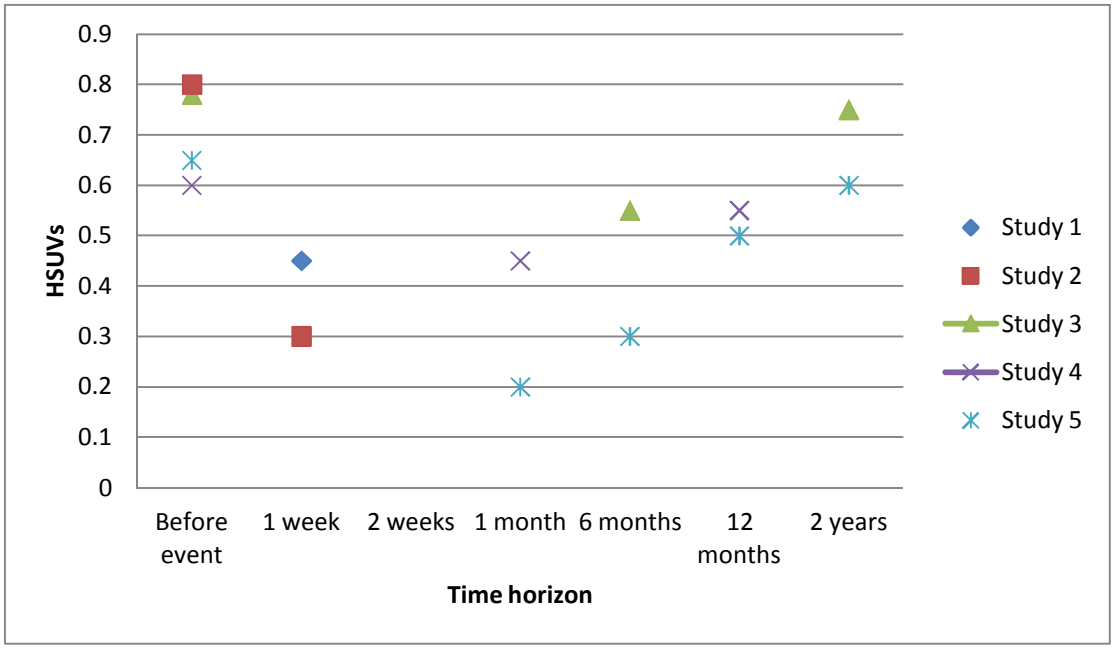
Every HSUV used in the decision model (either from a single study or a synthesised value) must be reported and full and explicit justification be given for its use. It is useful to structure this information by each health state requiring a HSUV in the decision model. The NICE template for manufacturer's submission provides a useful table by which to present such information (Table B15, section 6.4.9, p.41). If HUSVs are synthesised across studies, this should be made clear and an account given of the individual studies used. In addition, as mentioned previously, section 6.4.9 might be a useful place to highlight any studies that were excluded and the reasons for doing so (see section 3.4.2).

**Table 6: Table B15 (section 6.4.9) [26]: Summary of quality-of-life values for cost-effectiveness analysis**

| State | Utility value | Confidence interval | Reference in submission | Justification |
|---|---|---|---|---|
| Health state 1 | HS1 | | | |
| Health state 2 | HS2 | | | |
| Etc. | … | | … | |
| Adverse event 1 | AE1 | | | |
| Adverse event 2 | AE2 | | | |

The osteoporosis case study also provided visual presentation of results in the form of tables and scatter plots. Plotting HSUVs from different studies on scatter plot is a good way of visually presenting the similarities and differences in HSUV values. This might be useful when presenting contextual data, for example in highlighting how values differed between different studies as a result of differences in the populations or HRQL instrument used.

**Figure 7: Visual presentation of HSUV data**




### *7.2.1. Synthesised HUSVs*

Perhaps owing to the fact that relatively few examples exist of synthesised HSUVs and that it is not a requirement of the NICE methods guide, the NICE template for manufacturer submissions does not provide a place for discussion of synthesised HSUVs. Where values have been synthesised, this should be stated, along with a description of the individual studies used to generate the synthesised values (as described in section 7.2 of this TSD), and the methods used to synthesise the values. In addition, where synthesised values are used, there needs to be a discussion of the level of heterogeneity between studies used to calculate the value. If synthesis was not performed, it may be useful to provide the reasons why it was not undertaken (for example large variety of HRQL instruments used to derive HSUVs).

## 7.3. QUALITY AND RELEVANCE ASSESSMENT

The current NICE methods guide does not include a requirement to provide quality and relevance assessment for HSUV reviews. Although section 6.4.6 in the manufacturer/sponsor's template requests several items of information relating to quality and relevance of HSUV studies (e.g. sample size, response rates, consistency with reference case, appropriateness of health states given condition and treatment pathway), there is no natural place for these items to be discussed in any detail within the template. However, discussing the quality and relevance of included studies that provide HSUVs should be considered as

good practice using the criteria provided in section 4 (box 3 and 4). This could be in the form of a narrative synthesis or data could be tabulated. Particular emphasis should be placed on how study weaknesses might affect the robustness of HSUVs and how relevance might affect the application of HSUVs in decision models.

Despite both case study reviews indicating that data were extracted in relation to the items listed in Box 3 and 4 in Section 4, discussion around quality and relevance was limited in the case studies. The only quality criteria considered by the osteoporosis review was loss-to follow up, whilst the breast cancer review did not explore the quality of studies. Neither review provided quality assessment information in tables. However, factors relating to relevance were covered in more detail by the reviews. The osteoporosis review discussed the impact of exclusion criteria in the studies (such as institutionalised adults, secondary osteoporosis and presence of co-morbidities) on being able to infer HSUVs beyond the study population. The breast cancer review briefly discussed the impact of age and who is valuing the health states (e.g. patient or clinician) on the study findings. In addition, both studies provided detailed descriptions of how the choice of valuation method strongly impacts on the HSUV.

### 7.4. MODIFICATIONS TO HSUVS AND SENSITIVITY ANALYSIS

Where HSUVs are adapted or modified for use in the decision model (as described in TSD12[29]), full reasons need to be provided for why this was undertaken (will be informed by the relevance of the data extracted to the decision model and NICE reference case) and the methods of how it was undertaken. Section 6.4.15 in the NICE template for manufacturer sponsor submissions covers this presentation. Finally, where a range of HSUVs have been used in sensitivity analyses in the decision model, this process needs to be described.

## 8. RECOMMENDATIONS

This TSD suggests methods for systematically identifying, selecting, reviewing and synthesising HSUV values. Below, we present recommendations for undertaking HSUV reviews including methods to systematically identify and select HSUVs for decision models for HTA submissions to NICE. The challenges in undertaking HSUVs reviews have been explored in this TSD at each stage of the review process and section 8.2 provides recommendations for future research.

## 8.1. RECOMMENDATION FOR UNDERTAKING HSUV REVIEWS

- The scope and identification of the evidence for HSUV reviews need to be kept broad initially
- Ideally use a variety of resources and methods to identify relevant studies e.g. electronic database searching, reference list checking, contact with experts etc.
- The scope of the review and inclusion/exclusion criteria will be refined during the stage of evidence selection according to the nature of the evidence base
- Selecting evidence for HSUV reviews is an iterative process and may involve preliminary data extraction of key characteristics (population details, approach used to describe the health state, elicitation technique). Based on that data, decisions can be made on how to amend and develop inclusion criteria further.
- Selection of included studies must be well justified and explicit. A record of reasons for study inclusion and exclusion (i.e. those studies identified as possible but ultimately excluded) must be kept.
- Criteria for quality and relevance have been suggested in this guide (see box 3 and 4)
- Where there is more than one set of values meeting the reference case and that are relevant to the model, then the final selection needs to be justified with sensitivity analyses using alternative values, and consideration given to pooling to improve precision of estimated HSUVs.
- Data presentation must include an account of the search undertaken to identify studies, characteristics of included studies, HSUVs used in the decision model (with justification), quality and relevance assessment and modifications made to values used in the model.

## 8.2. RECOMMENDATIONS FOR FUTURE RESEARCH

### *8.2.1. Database/Register of HSUVs*

A programme of work is required to identify and collate reviews of HSUVs. For example our case studies provide HUSVs for breast cancer and osteoporosis-related conditions (e.g. hip fracture) and submissions to NICE will include reviews of HSUVs in a range of topic areas. The results of these reviews should be publically available in a user friendly form. Where

gaps are identified or evidence is lacking, reviews should be commissioned, and these could be particularly directed towards those disease areas where investing in HSUV reviews would be valuable because of their wide applicability to many HTAs e.g. cancer, cardiovascular disease and diabetes. Secondly, ensuring access to these reviews and HSUVs by some form of central resource, in a similar manner to the Cochrane Library that allows free and timely access to reviews of clinical effectiveness, would allow a wider sharing of these resources. Similarly, a register of primary studies would also be useful, analogous to the CENTRAL database at the Cochrane Library

### *8.2.2. Identifying HSUV data via literature searching*

As discussed in section 3, standard methods for identifying HSUVs do not exist. Section 3 refers to one widely used but unvalidated quality of life filter,[32] and provides some of the most commonly used MEDLINE (MeSH) and Embase (EMTREE) terms. This stage of the HSUV review process requires further research; however simply creating a validated filter may not be useful since it is unlikely that this will solve the problems in searching electronic databases with acceptable sensitivity and specificity for HSUVs, unless problems in indexing and abstract-level reporting improve. Identifying studies by database searching is as difficult due to the tendency for unmanageable result sets to be retrieved by the necessity of keeping the approach broad to begin with. These problems also have implications for the data selection stage when citations are sifted according to inclusion criteria. If reporting of HSUVs in abstracts is poor this will result in relevant studies being missed or excessive and time-intensive examination of full-texts.

It was interesting to note that in the osteoporosis review case study, the majority of references were identified by means other than electronic searching. It is also worth considering that the time and resources that go into clinical effectiveness reviews are extensive, and is it feasible for HSUV reviews and more pertinently an appropriate use of time in HSUV reviews? Further research directed towards the type of search approach in relation to the scope of HSUV review, i.e. exploration of the sensitivity and specificity of different search approaches such as the supplementary techniques is required.

### *8.2.3. Methods of synthesis*

While the synthesis of clinical parameters is a well developed area of research, there has been little research into the synthesis of HSUVs. The problem until recently has been in obtaining any relevant HSUVs. With a growing literature of values there are increasingly situations where an analyst will have a number of values to choose from. To make the best use of such evidence some form of synthesis will maximise the reliability and precision of any estimates. Methods need to be further explored those being applied in clinical reviews such as the use of Bayesian approaches.

### *8.2.4. Time/resource issues*

Conducting systematic reviews is time and resource intensive, and the guidance in NICE methods guide not does stipulate that a systematic review of HSUVs be undertaken. Nevertheless identification and selection of HSUVs must be systematic and transparent, and this in itself involves considerable resources. Furthermore, where 'added value processes' are used for HSUV reviews such as quality assessment and data synthesis, this further increases the resources required. Typically a systematic review of clinical effectiveness involves a dedicated information specialist and one to two systematic reviewers. In theory, the same model should be applied to HSUV reviews, however in the context of technology appraisals this is likely to be unfeasible and it is not strictly necessary. Investigating ways that information specialists and systematic reviewers could aid with the HSUV review in assessment of relevance of material might aid this process. However, ideally moving to a model whereby HSUV data can be identified through existing databases or registers of HSUVs is the longer term solution, as outlined in section 8.2.1.

### *8.2.5. Quality and relevant assessment*

We have proposed criteria for assessing the quality and relevance of HSUV studies in section 4. However, these criteria need to be tested and if necessary amended or added to.

## 9. SUMMARY

HSUVs are important parameters in decision models, and thus the methods to identify, review and select appropriate values must be considered carefully. NICE requires evidence that HSUV estimates from published literature have been identified and selected systematically.[4] The principles of systematic reviewing for clinical effectiveness reviews can

inform some aspects of how to identify and select utilities systematically, but there are unique issues to be explored in the scoping and identification of evidence of HSUVs reviews. The process of identifying and selection evidence differs in that often a sequence of searches may be required, rather than one literature search. Study selection also informs the inclusion criteria in terms of refining the type of HSUV data to be included according to the evidence available. The process of evidence assessment involves both quality and relevance assessment. This TSD provides guidance on how to scope, search, select, quality and relevance assess, data extract and present data. The final selection of values used in the model needs to be justified and sensitivity analysis undertaken of alternative possible values.

## 10. REFERENCES


1. Paisley, S., Booth, A., Mensinkai, S.. Health-related quality of life studies. *Etext on Health Technology Assessment (HTA) Information Resources.* US National Library of Medicine (NLM) National Information Center on Health Services Research and Healthcare Technology (NICHSR); Bethesda: 2005.

2. National Institute for Health and Clinical Excellence (NICE). Guide to the single technology appraisal process. 2009. NICE.

3. Torrance, G.W. Measurement of health state utilities for economic appraisal: a review. *Journal of Health Economics* 1986; 5(1):1-30.

4. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. 2008. NICE.

5. Tosh, J., Longworth, L., George, E. Utility values in NICE Technology Assessments . *Value in Health* 2011; 14(1).

6. Centre for Reviews and Dissemination. CRD's guidance for undertaking reviews in healthcare. 3rd ed. Centre for Reviews and Dissemination, York; 2009.

7. Higgins, J.P.T., Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 ed. Cochrane Collaboration, Oxford; 2008.

8. Peasgood, T., Herrmann, K., Kanis, J.A., Brazier, J. An updated systematic review of Health State Utility Values for osteoporosis related conditions. *Osteoporosis International* 2009; 20(6):853-868.

9. Peasgood, T., Ward, S.E., Brazier, J. A review and meta analysis of health state utility values in breast cancer. *Expert Review of Pharmacoeconomics and Outcomes Research* 2010; 10(5):553-566.

10. Kanis, J.A., Brazier, J., Stevenson, M.D., Calvert, N.W., Lloyd-Jones, M. Treatment of established osteoporosis: as systematic review and cost utility analysis. *Health Technology Assessment* 2002; 6(29):1-146.

11. Stevenson, M.D., Brazier, J., Calvert, N.W., Lloyd-Jones, M., Oakley, J., Kanis, J.A. Description of an individual patient methodology for calculating the cost-effectiveness of treatments for osteoporosis in women. *Journal of the Operational Research Society* 2005; 56(2):214-221.

12. Brazier, J., Green, C., Kanis, J.A. A systematic review of Health State Utility Values of Osteoporosis-related conditions. *Osteoporosis International* 2002; 13(10):768-776.

13. Newcomb, P.A., Trentham-Dietz, A., Hampton, J.M. Bisphosphonates for osteoporosis treatment are associated with reduced breast cancer risk. *British Journal of Cancer* 2010; 102(5):799-802.

14. Vestergaard, P., Schwartz, K., Pinholt, E.M., Rejnmark, L., Mosekilde, L. Risk of atrial fibrillation associated with use of bisphosphonates and other drugs against osteoporosis: a cohort study. *Calcified Tissue International* 2010; 865(5):335-342.

15. Jeffcoat, M.K., Cizza, G., Shih, W.J., Gneco, R., Lombardi, A. Efficacy of bisphosphonates for the control of alveolar bone loss in periodontitis. *Journal of the International Academy of Periodontology* 2007; 9(3):70-76.

16. 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

17. 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

18. Centre for the Evaluation of Value and Risk in Health. CEA Registry. 2010; available from https://research.tufts-nemc.org/cear/Default.aspx

19. Centre for Reviews and Dissemination. HTA. 2010; available from www.crd.york.ac.uk/crdweb.

20. Centre for Reviews and Dissemination. NHS Economic Evaluation Database. 2010; available from www.crd.york.ac.uk/crdweb.

21. Wiley Interscience. *HEED: Health Economics Evaluation Database*. 2010; available from http://heed.wiley.com.

22. EuroQol Group. *EQ-5D: a standardised instrument for use as a measure of health outcome*. 2010; available from www.euroqol.org

23. MAPI Research Institute. *MAPI Institute*. 2010; available from www.mapi-institute.com/home.

24. RePEc. Research Papers in Economics (RePEc) . 2010; available from http://repec.org/

25. MAPI Institute. Patient-reported outcome and quality of life instruments database (PROQOLID) . 2010; available from http://www.proqolid.org/

26. National Institute for Health and Clinical Excellence (NICE). Specification for manufacturer/sponsor submission of evidence. 2009.

27. Critical Appraisal Skills Programme (CASP). 10 questions to help you make sense of reviews. 2006;

28. Czoski-Murray, C., DeNigris, E., Brazier, J., Walters, S. A prospective controlled study of the costs and health realted quality of life following hip fracture. 2007. School of Health and Related Research, University of Sheffield.

29. 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

30. Tengs, T.O., Lin, T.H. A meta-analysis of utility estimates for HIV/AIDS. *Medical Decision Making* 2002; 22(6):475-481.

31. Post, P.N., Stiggelbout, A.M., Wakker, P.P. The utility of health states after stroke: a systematic review of the literature. *Stroke* 2001; 32(6):1425-1429.

32. InterTASC Information Specialists' Sub-Group. *Search Filter Resource: quality of life sample search filter*. 2010; available from www.york.ac.uk/inst/crd/intertasc/qol1.htm.

# APPENDIX

**Appendix A. 1: Medline search strategy for osteoporosis case study**

| No | Request |
|---|---|
| 1 | *Osteoporosis/ |
| 2 | *Bone Diseases, Metabolic/ |
| 3 | osteoporo$.ti. |
| 4 | 1 or 2 or 3 |
| 5 | (bone adj6 densit$).ti. |
| 6 | *Bone Density/ |
| 7 | (bone or bones).ti. |
| 8 | *densitometry/ |
| 9 | *Tomography, X-Ray Computed/ |
| 10 | densit$.ti. |
| 11 | 9 and 10 |
| 12 | 8 or 11 |
| 13 | 7 and 12 |
| 14 | 5 or 6 or 13 |
| 15 | *Colles' Fracture/ |
| 16 | *hip fractures/ |
| 17 | *Spinal Fractures/ |
| 18 | 15 or 16 or 17 |
| 19 | *Fractures, Bone/ |
| 20 | fractur$.ti. |
| 21 | 18 or 19 or 20 |
| 22 | colles$.ti. |
| 23 | (hip or hips).ti. |
| 24 | (femur adj6 neck).ti. |
| 25 | (femoral adj6 neck).ti. |
| 26 | (spine or spinal).ti. |
| 27 | vertebra$.ti. |

| | |
|---|---|
| 28 | *Lumbar Vertebrae/ |
| 29 | 22 or 23 or 24 or 25 or 26 or 27 or 28 |
| 30 | 21 and 29 |
| 31 | *Estrogen Replacement Therapy/ |
| 32 | estrogen replacement therapy.ti. |
| 33 | oestrogen replacement therapy.ti. |
| 34 | hormone replacement therapy.ti. |
| 35 | ert.ti. |
| 36 | ort.ti. |
| 37 | hrt.ti. |
| 38 | 31 or 32 or 33 or 34 or 35 or 36 or 37 |
| 39 | *menopause/ |
| 40 | *Climacteric/ |
| 41 | menopaus$.ti. |
| 42 | postmenopaus$.ti. |
| 43 | climacteric.ti. |
| 44 | 39 or 40 or 41 or 42 or 43 |
| 45 | 38 or 44 |
| 46 | 30 and 45 |
| 47 | 4 or 14 or 46 |
| 48 | exp quality of life/ |
| 49 | quality of life.tw. |
| 50 | life quality.tw. |
| 51 | hql.tw. |
| 52 | (sf 36 or sf36 or sf thirtysix or sf thirty six or short form 36 or short form thirty six or short form thirty-six or short form 36).tw. |
| 53 | qol.tw. |
| 54 | (euroqol or eq5d or eq 5d).tw. |
| 55 | qaly$.tw. |
| 56 | quality adjusted life year$.tw. |
| 57 | hye$.tw. |
| 58 | health$ year$ equivalent$.tw. |

| | |
|---|---|
| 59 | health utilit$.tw. |
| 60 | hui.tw. |
| 61 | quality of wellbeing$.tw. |
| 62 | quality of well being.tw. |
| 63 | qwb.tw. |
| 64 | (qald$ or qale$ or qtime$).tw. |
| 65 | standard gamble$.tw. |
| 66 | time trade off.tw. |
| 67 | time tradeoff.tw. |
| 68 | tto.tw. |
| 69 | visual analog$ scale$.tw. |
| 70 | discrete choice experiment$.tw. |
| 71 | (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or short form six or short form six).tw. |
| 72 | health state$ utilit$.tw. |
| 73 | health state$ value$.tw. |
| 74 | health state$ preference$.tw. |
| 75 | or/48-74 |
| 76 | letter.pt. |
| 77 | editorial.pt. |
| 78 | comment.pt. |
| 79 | or/76-78 |
| 80 | 75 not 79 |
| 81 | 47 and 80 |
| 82 | limit 81 to yr="2000 - 2007" |

**Appendix A. 2: CASP Checklist**[v]

| **CASP checklist for reviews- 10 questions**[27] |
|---|
| **Did the review ask a clearly-focused question?**<br>*Consider if the question is 'focused' in terms of:*<br>*– the population studied*<br>*– the intervention given or exposure*<br>*– the outcomes considered* |
| **Did the review include the right type of study?**<br>*Consider if the included studies:*<br>*– address the review's question*<br>*– have an appropriate study design* |
| **Did the reviewers try to identify all relevant studies?**<br>*– which bibliographic databases were used*<br>*– if there was follow-up from reference lists*<br>*– if there was personal contact with experts*<br>*– if the reviewers searched for unpublished studies*<br>*– if the reviewers searched for non-English-language studies* |
| **Did the reviewers assess the quality of the included studies?**<br>*Consider:*<br>*– if a clear, pre-determined strategy was used to determine which studies were included. Look for:*<br>*– a scoring system*<br>*– more than one assessor* |
| **If the results of the studies have been combined, was it reasonable to do so?**<br>*– the results of each study are clearly displayed*<br>*– the results were similar from study to study (look for tests of heterogeneity)*<br>*- the reasons for any variations in results are discussed* |
| **How are the results presented and what is the main result?**<br>*how the results are expressed (e.g. odds ratio, relative risk, etc.)*<br>*– how large this size of result is and how meaningful it is*<br>*– how you would sum up the bottom-line result of the review in one sentence* |
| **How precise are these results?**<br>*Consider:*<br>*– if a confidence interval were reported. Would your decision about whether or not to use this intervention be the same at the upper confidence limit as at the lower confidence limit?*<br>*– if a p-value is reported where confidence intervals are unavailable* |
| **Can the results be applied to the local** population**?**<br>*Consider whether:*<br>*– the population sample covered by the review could be different from your population in ways that would produce different results*<br>*– your local setting differs much from that of the review*<br>*– you can provide the same intervention in your setting* |

[v] The CASP Systematic Reviews checklist has been reproduced in Appendix x and adapted for use in HSUV reviews (in section 4 of the TSD) with the kind permission of CASP at Solutions for Public Health. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the Public Health Resource Unit. © Public Health Resource Unit, England 2006

| |
|---|
| **Were all important outcomes considered?**<br>*Consider outcomes from the point of view of the:*<br>*– individual*<br>*– policy makers and professionals*<br>*– family/carers*<br>*– wider community* |
| **Should policy or practice change as a result of the evidence contained in this review?**<br>*Consider:*<br>*– whether any benefit reported outweighs any harm and/or cost. If this information is not reported can it be filled in from elsewhere?* |

**Appendix A. 3: Sample data extraction form**

General information:
Name of Data extractor..... Date of data extraction: .....

| Study Ref ID | Study details- author, title, year | Country of respondents | Study design | Inclusion and exclusion criteria | Disease-related health state, including time horizon | Participant characteristics where HRQL change measured | Participant characteristics used in valuation of HRQL change | Details of health state description system (if applicable) | Method of elicitation of HSUVs - how and who? | Valuation technique e.g. SG, TTO, VAS |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Including:<br>Age<br>Sex<br>Disease severity<br>Any other relevant characteristics e.g. Setting | | Lower and upper bound- death or worst possible health, perfect health or normal health | e.g. how- vignettes/scenarios, health state descriptive system (e.g. EQ-5D), direct measurement<br><br>e.g. who- patient, public, clinician | |

| Respondent selection and recruitment | Response rates | Reasons for lost to follow-up | Any other potential problems with the study | HSUV value descriptive statistics: per subgroup | | | |
|---|---|---|---|---|---|---|---|
| | | | | Sample size/No. of respondents | Means (SD) | Medians | Range |
| | | | | | | | |

**Appendix A. 4: Headings of the NICE manufacturer/sponsor submission template relating to measurement and valuation of health effects (p. 38-42)**

## 6.4 Measurement and valuation of health effects

### Patient experience

**6.4.1** Please outline the aspects of the condition that most affect patients' quality of life.

**6.4.2** Please describe how a patient's HRQL is likely to change over the course of the condition.

### HRQL data derived from clinical trials

**6.4.3** If HRQL data were collected in the clinical trials identified in section 5 (Clinical evidence), please comment on whether the HRQL data are consistent with the reference case. The following are suggested elements for consideration, but the list is not exhaustive.

- Method of elicitation.
- Method of valuation.
- Point when measurements were made.
- Consistency with reference case.
- Appropriateness for cost-effectiveness analysis.
- Results with confidence intervals.

### Mapping

**6.4.4** If mapping was used to transform any of the utilities or quality-of-life data in clinical trials, please provide the following information.

- Which tool was mapped from and onto what other tool? For example, SF-36 to EQ-5D.
- Details of the methodology used.
- Details of validation of the mapping technique.

### HRQL studies

**6.4.5** Please provide a systematic search of HRQL data. Consider published and unpublished studies, including any original research commissioned for this technology. Provide the rationale for terms used in the search strategy and any inclusion and exclusion criteria used. The search strategy used should be provided in section 9.12, appendix 12.

**6.4.6** Provide details of the studies in which HRQL is measured. Include the following, but note that the list is not exhaustive.

- Population in which health effects were measured.
- Information on recruitment.
- Interventions and comparators.
- Sample size.
- Response rates.
- Description of health states.
- Adverse events.
- Appropriateness of health states given condition and treatment pathway.
- Method of elicitation.
- Method of valuation.
- Mapping.

- Uncertainty around values.
- Consistency with reference case.
- Appropriateness for cost-effectiveness analysis.
- Results with confidence intervals.
- Appropriateness of the study for cost-effectiveness analysis.

**6.4.7** Please highlight any key differences between the values derived from the literature search and those reported in or mapped from the clinical trials.

**Adverse events**

**6.4.8** Please describe how adverse events have an impact on HRQL.

**Quality-of-life data used in cost-effectiveness analysis**

**6.4.9** Please summarise the values you have chosen for your cost-effectiveness analysis in the following table, referencing values obtained in sections 6.4.3 to 6.4.8. Justify the choice of utility values, giving consideration to the reference case.

**Table 7: Table B1 Summary of quality-of-life values for cost-effectiveness analysis**

| State | Utility value | Confidence interval | Reference in submission | Justification |
|---|---|---|---|---|
| Health state 1 | HS1 | | | |
| Health state 2 | HS2 | | | |
| Etc. | … | | … | |
| Adverse event 1 | AE1 | | | |
| Adverse event 2 | AE2 | | | |

**6.4.10** If clinical experts assessed the applicability of values available or estimated any values, please provide the following details :

- the criteria for selecting the experts
- the number of experts approached
- the number of experts who participated
- declaration of potential conflict(s) of interest from each expert or medical speciality whose opinion was sought
- the background information provided and its consistency with the totality of the evidence provided in the submission
- the method used to collect the opinions
- the medium used to collect opinions (for example, was information gathered by direct interview, telephone interview or self-administered questionnaire?)
- the questions asked
- whether iteration was used in the collation of opinions and if so, how it was used (for example, the Delphi technique).

**6.4.11** Please define what a patient experiences in the health states in terms of HRQL. Is it constant or does it cover potential variances?

**6.4.13** If appropriate, what was the baseline quality of life assumed in the analysis if different from health states? Were quality-of-life events taken from this baseline?

**6.4.14** Please clarify whether HRQL is assumed to be constant over time. If not, provide details of how HRQL changes with time.

**6.4.15** Have the values in sections 6.4.3 to 6.4.8 been amended? If so, please describe how and why they have been altered and the methodology.