
Article:
Ara, R.M. orcid.org/0000-0002-7920-1707, Brazier, J., Peasgood, T. et al. (1 more author) (2017) The Identification, Review and Synthesis of Health State Utility Values from the Literature. PharmacoEconomics, 35 (Suppl 1). pp. 43-55. ISSN 1170-7690

https://doi.org/10.1007/s40273-017-0547-8

THE IDENTIFICATION, REVIEW AND SYNTHESIS OF HSUVs FROM THE LITERATURE

Running header: Identifying health state utility values

Authors: Roberta Ara, MSc[1]; John Brazier, PhD[1]; Tessa Peasgood, PhD[1]; Suzy Paisley, PhD[1]

[1]University of Sheffield

ScHARR

Regent Street

University of Sheffield

30 Regent Street

Sheffield

S1 4DA

Corresponding author: Tessa Peasgood (t.peasgood@sheffield.ac.uk)

Abstract

Systematic literature reviews of health related quality of life (HRQoL) evidence that are to inform economic models can be challenging due to the volume of hits identified in searches using generic terms for HRQoL. Nevertheless, a robust review of the literature is required to ensure that the health state utility values (HSUV) used in the economic model are the most appropriate available. This article provides a synopsis of literature relating to identifying, reviewing and synthesising HSUVs.

The process begins with scoping the needs of the economic model including the definitions of health states and the requirements of any reimbursement agencies. A sequence of searches may be required as the economic model evolves. The terminology used for HRQoL measures may be problematic and as there is no robust HRQoL filter (equivalent to that applied for RCTs), sifting the results of sensitive searches can be resource intensive. Alternative approaches such as forward and backward citation searches may reduce the resources required while maintaining the integrity of the search.

Any included studies should be assessed in terms of quality using a recommended checklist and insufficient detail in the primary studies should be noted as a short-coming in this exercise. Subject to homogeneity (similar populations, same measure and preference weights) evidence can be pooled in some way, although methodological research into the appropriateness of alternative techniques for meta-analysis is in its infancy. Reporting standards are key and as a minimum should include details on searches, inclusion/exclusion criteria (together with rationale for exclusion at each stage), assessment of quality and relevance of included studies, and justification for the choice of final HSUVs.

Key points for decision makers

- A variety of resources and methods should be used to identify studies including relevant electronic databases, reference list searches, key author and citations searches and contact with experts.
- Identification of evidence may be an iterative process and should be informed by the decision model and requirements of relevant reimbursement agencies.
- As searches and inclusion criteria are likely to evolve overtime, records of searches and exclusion rationale should be well documented and final selection should be justified.

1. Introduction

Health state utility values (HSUVs) are important parameters in economic models and with an ever increasing evidence base it is becoming increasingly important to demonstrate that the HSUVs are identified and selected using a systematic and transparent method. However, modelling reports frequently provide very sparse descriptions of the HSUVs used and rarely document the source or justify the values selected[1].

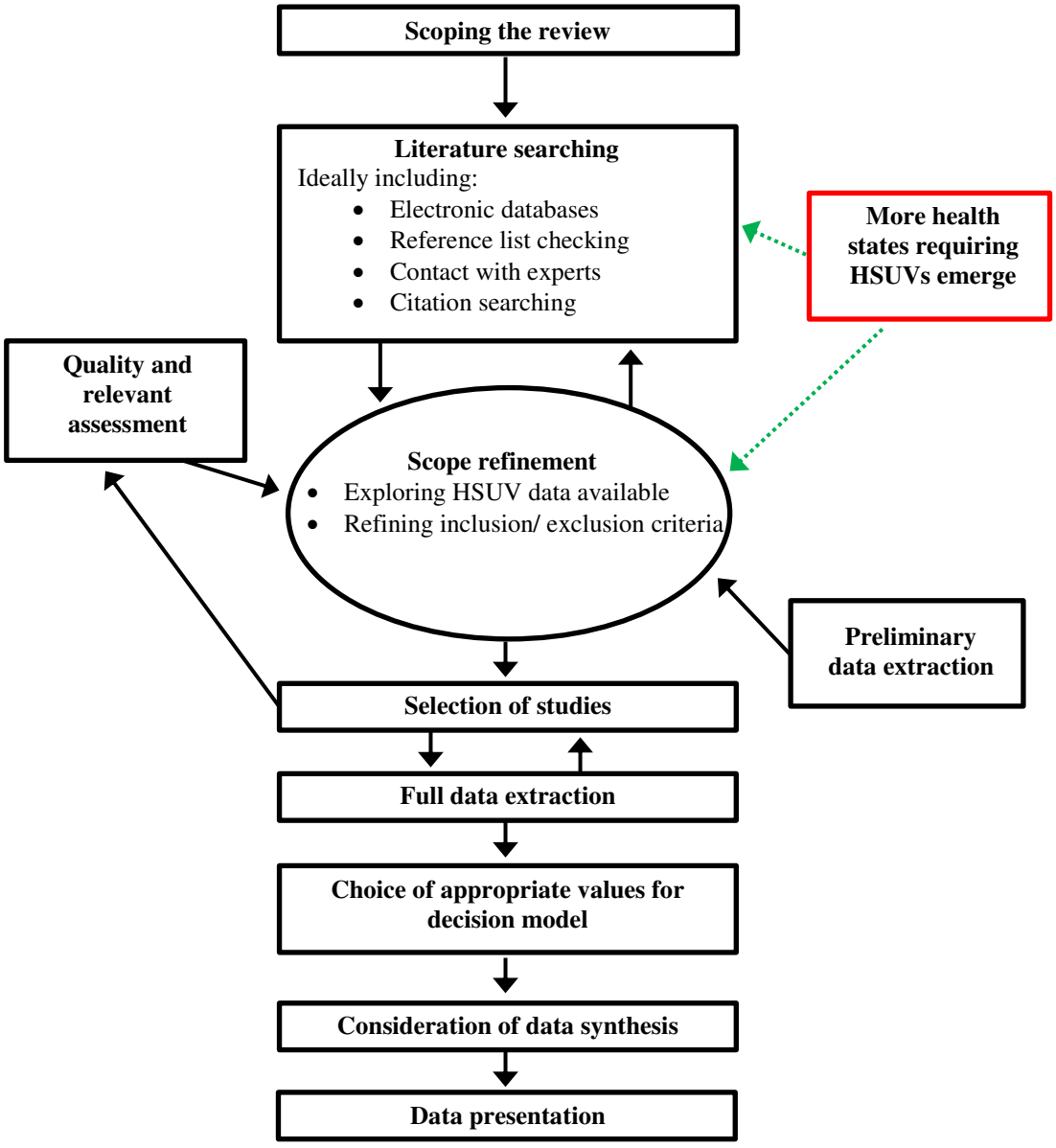
This article provides an overview of the current literature relating to identifying and reviewing HSUVs from the literature, together with details on when and how to synthesise the evidence. A case-study of a review in osteoporosis related conditions is used for demonstration purposes. The study aimed to systematically review published HSUVs for osteoporotic fractures (hip, clinical, and morphometric vertebral, wrist, shoulder, and the interaction of several fractures) to determine the associated loss of utility for use in future decision analytic models [2,3]. Full details of this study are reported elsewhere [2,4,5]. Checklists and recommendations for reviewing published HSUV evidence are provided. Terminology used in this article includes:

- Sensitivity – the ability of the literature search to find all relevant material
- Precision – the ability of the literature search to reject irrelevant material

2. Scoping the review

As the aim of the review is to populate the health states within a decision analytic model, prior to embarking on the review, it is essential to fully scope the needs of the model as far as is possible to determine the HSUVs required. This includes the health states defined within the model and the requirements of any reimbursement agencies that will use the results generated to inform policy decision making. Whilst the scoping stage will arrive at a clear understanding of the required HSUVs at the outset of the review, a degree of iteration and scope refinement is possible as the model development process progresses. **Figure 1** provides details on the possible stages in the systematic reviewing process.

Figure 1 Reviewing HSUVs



Source: Reprinted from Papaioannou et al. [5], page 687, with permission from Elsevier

2.1 Health states within the model

Decision analytic models describe the clinical pathway of a condition through the use of discrete health states. These models can become quite complex with multiple health states representing: the progression of the underlying condition (such as decreases in function and increases in pain levels as observed in arthritic conditions), condition related comorbidities (such as onset of cardiovascular disease in diabetes), discrete events (such as different fractures in osteoporosis), adverse events associated with the interventions (such as gastrointestinal problems for people on

long-term steroid use), or conversely, additional benefits beyond the condition of interest (for example, bisphosphonates provided to treat osteoporosis may also reduce the risk of breast cancer, and bone loss in periodontal disease). **Figure 2** provides the possible health states in a model in osteoporosis. The time horizon used can affect requirements as can be seen in **Figure 2** where the HSUV immediately post fracture may differ from the HSUV at one year post fracture. Additional HSUVs may also be required as the economic model evolves over time.

Figure 2 HSUV data needs in the osteoporosis case-study

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

Source: Reprinted from Papaioannou et al. [5], page 689, with permission from Elsevier

2.2 The type of evidence required

The type of evidence required may be dictated by the decision maker/reimbursement authority [6]. For example, while the National Institute for Health and Care Excellence NICE in the UK states a preference for the EQ-5D evidence where possible, other agencies do not place similar restrictions

on the type of evidence and ask that the selection of evidence used in submission is 'justified' (e.g. Canadian Agency for Drug and Technologies in Health (CADTH) in Canada) [7].

The hits obtained from searches for HSUVs can be notoriously large as the evidence base is not restricted to randomised control trials (RCTs). HSUVs can be obtained from various measures (e.g. condition-specific preference-based measures (CSPBMs) or generic preference-based measures (GPBMs)) [8,9] and restricting the searches to a particular measure may help with managing the number of possible hits from searches. However, introducing a filter (such as EQ-5D) too early in the scoping phase may limit the usefulness of the search resulting in no studies being selected for review. Consequently, if there is a specific requirement (such as EQ-5D in the first instance), a scoping search to gauge the size of the likely evidence is recommended. If the initial scoping search identifies no evidence which satisfies the requirements of both the reimbursement authority and the model, then a broader search strategy will be required (perhaps expanding the GPBM beyond the initial preferred measure). It should be borne in mind that it may be possible to map from the measure collected in a clinical efficacy study to the preference-based measure required by the agency [10].

2.3 Appropriateness of existing systematic reviews

Existing systematic reviews, such as those published as journal articles, or reviews in previous Health Technology Assessment (HTA) submissions can provide the required HSUVs, or they can be used as the starting point of a future search strategy. Any scoping searches should be sufficiently sensitive to identify this form of evidence. However, prior to using the HSUVs reported within the review, or using the search strategies used in the review to inform the new searches, the review needs to be assessed in terms of quality and in particular: are the identification and selection methods systematic and transparent and has the evidence reported been appraised satisfactorily. An adapted version of the CASP checklist (**Table 1**) [is recommended in the latest NICE Technical Support Document (TSD) and the unmodified version is reproduced in full in the TSD [4].

Table 1 Quality assessment checklist for systematic reviews of HSUVs

	Tick as appropriate			
	Yes	**No**	**Partly**	**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 state (ideally patients)* *Population valuing the change in HRQoL (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: *Addressed the review's question* *Are appropriate studies*				
3. Did the review 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?* Ideally, but not mandatory, consider that the search methods should involve: *Personal contact with experts* *Search for unpublished studies* *Citation and author searches*				
4. Did the review assess the quality of the included studies? Consider the: *Sample size* *Respondent selection and recruitment* *Inclusion/exclusion criteria* *Response rates to instrument used* *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?*				

	Tick as appropriate			
	Yes	No	Partly	Can't tell
5. Did the review assess the relevance of the included studies to the review question? *Population describing the health state (ideally patients)* *Population valuing the change in HRQoL (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 variation 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 full justification of why studies were included?* *How are the results expressed (descriptive statistics or coefficients of a model)*				
8. How precise are the results? Consider: *If confidence intervals (CI) were reported. Would your decision about whether or not to use this INTERVENTION be the same at the upper CI as at the lower CI?* *If a p-value is reported where Cis are unavailable*				
9. Can the HUSVs be used for the health states in your decision model? Consider: *How relevant the population describing the health state is to the health state in the decision model* *Have all subgroups been considered e.g. age, disease severity, setting* *Do the HSUVs match the REQUIRED reference case* *How do the results need to be modified for the decision model?*				

Source: Adapted from CASP[11] and Papaioannou et al.[5], page 694, with permission from Elsevier **Key**: CI – confidence interval; HRQoL – health related quality of life; HSUV – health state utility value; TTO – time trade off.

3. Literature searching

Empirical evidence describing optimal approaches when searching for HSUVs is sparse, and methodological guidance used for reviews of clinical effects do not generalise to HSUVs. In addition, while there are several filters that can be used to limit the type of study design (e.g. RCT), there is no known validated methodological search filter for HSUV evidence although some exploratory work is being undertaken [12]. This absence of an established filter is partly due to the variety of potential sources (in addition to RCTs, HSUVs may also be reported in observational studies, cost-effectiveness analyses, HTA reports and industry submission documents). The ideal search would take a sensitive approach using a comprehensive search strategy encompassing a variety of sources (multiple electronic databases) supplemented by additional techniques such as citation searches and reference list checks. However, sensitive searches of this nature can be resource intensive in terms of sifting the potential hits identified. An alternative, which retains the systematic and replicable nature of the search, may be to restrict the main search in the expectation that the supplementary search techniques are likely to detect most unidentified evidence.

Searches for HSUVs are likely to involve a sequence of searches as the model evolves (the numbers or definitions of the health states within the model increase or change respectively), and the results of the initial literature searches identify gaps in the required evidence. Consequently, records of the individual searches and the selections from the corresponding results should be maintained to increase transparency of the process. Three core issues (search terms, where to search, and supplementary searches) that arise when searching for HSUVs are discussed below.

3.1 Search terms

While the terminology for the health states within the decision model is relatively straightforward (examples from the 28 possible health states in **Figure 2**: established osteoporosis with vertebral deformity, fracture of the hip, fracture of the shoulder, fracture of the wrist, etc.) the terminology for the health related quality of life (HRQoL) measures (i.e. the particular HSUVs required), can be more problematic in terms of both subject headings and the presence of relevant words in titles and abstracts.

Subject headings: The common GPBMs (i.e. EQ-5D, SF-6D, and HUI) do not have dedicated thesauri terms in Medline (MeSH) and Embase (EMTREE), and relevant studies can be indexed under broader concepts not obviously applicable to identifying HSUVs such as ‘quality of life’ or ‘quality adjusted life year’ (QALY). For example, when using HSUV-related free-text terms to identify a cross-sectional

sample of 300 records from Medline, the following indexing terms (in order of frequency) were identified: quality of life, questionnaires, psychology (subheading), health status, health status indicators, activities of daily living, health surveys, quality-adjusted-life-years, treatment outcome, psychometrics [13]. In the osteoporosis case-study review, of the 28 studies included, 24 were indexed on Medline, and of these 24 studies all but one were assigned 'Quality of life' as the MeSH term. While eight of the 24 studies were also assigned the MeSH term 'questionnaires', the majority of the other MeSH terms listed previously were not assigned to the studies in the case-study.

Although this example may suggest that the use of 'Quality of life' or 'Questionnaires' as MeSH terms in reviews for HSUVs may achieve a highly sensitive retrieval of relevant studies this likely to be coupled with very low precision, increasing the number of irrelevant studies retrieved substantially. There is always a trade-off between increasing the sensitivity or precision of searches, and this is particularly relevant for HSUVs where there are no useful validated terms that can be applied to filter results for studies reporting HSUVs.

Free-text searching: Appropriate free text terms are encompassed under three broad categories: general terms such as QALY and HSUV; instrument specific terms such as EQ-5D and SF-6D, and terms describing the methods of elicitation such as time trade off and standard gamble. Terms may be spelled differently, or the full names of measures may be used as opposed to the normal acronym (e.g. EQ-5D, eq5d, euoroqol, euro qol etc.) and exemplars of relevant free-text terms used in electronic databases are provided in **Figure 3**. Searches using free-text terms rely on their presence in the title or abstract but as HSUVs are frequently reported as secondary outcomes, they may not be mentioned in either and relevant studies will be missed. In the case-study in osteoporosis, of the 24 studies indexed on Medline, 17 included 'quality of life' in the title and 5 contained 'quality of life' in the abstract. Twenty two of the 24 included free text terms in either their title or abstract that related to the specific measure used to collect HSUVs.

Figure 3 Frequently used free-text terms for HSUVs when searching electronic databases

quality adjusted life
quality-adjusted-life (note not all databases can hope with hyphens)
(qaly$ or qald$ or qale$ or qtime$)
disability adjusted life
daly$
health$ year$ equivalent$
hsuv$
utilit$
(hql or hqol or h qol or hrqol or hr qol)
(hye or hyes)
disutili$
disbenefit$
preference-based measure$
generic measure$
gpbm$
preference elicit$
quality adj2 wellbeing
qwb

(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)
(hui or hui1 or hui2 or hui3)

Discrete choice
Dce$
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

Source: adapted from Papaiannou et al. [5], page 691, with permission from Elsevier

3.2 Sources to search

The main source for identifying studies containing HSUVs is electronic databases. In addition to the main health-related databases and conference proceedings, specialist health economic resources can provide additional relevant studies (**Table 2**).

Specialist sources include the Tufts Cost-effectiveness Analysis (CEA) Registry, the Health Technology Assessment database (HTA) and the NHS Economic Evaluation Database (NHS EED). The latter is no longer being updated but nonetheless remains a useful resource. An additional resource is ScHARRHUD (http://www.scharrhud.org), a free, searchable web-based database developed and managed by ScHARR at the University of Sheffield. The database holds bibliographic details of studies reporting HSUVs. ScHARRHUD is searchable by the names of the health state valuation instruments used in each study. In providing this level of indexing, ScHARRHUD aims to make the retrieval of health utilities evidence both more efficient and effective. Whilst the focus of ScHARRHUD is on GPBMs, such as the EQ-5D, the database extracts and indexes the names of all instruments used in studies. Whilst none of these specialist sources claim to be comprehensive, they do, collectively, provide a useful focus on the retrieval of HSUV evidence.

Table 2 Sources searched to identify studies including HSUVs

Electronic Databases:
Main health-related databases MEDLINE EMBASE CENTRAL
Conference Proceedings ISI Proceedings ISPOR ISOQoL Google Scholar
Additional specialist health economic resources Tufts Cost-effectiveness Analysis Registry (former Harvard Cost-Effectiveness Analysis database) http://healtheconomics.tuftsmedicalcenter.org/cear4/Home.aspx Centre for Reviews and Dissemination databases: NHSEED and HTA https://www.crd.york.ac.uk/CRDWeb/ The EQ-5D Website / other instrument sites http://www.euroqol.org/about-eq-5d.html MAPI Research Trust (including the PROQOLID database) http://mapi-trust.org/ Submissions to NICE or other health care decision agencies (e.g. Canadian Agency for Drugs and Technologies in Health (CADTH), Haute Autorite de Sante (HAS), Pharmaceutical Benefits Advisory Committee (PBAC),) ScHARR Health Utilities Database (ScHARRHUD) http://scharrhud.org/
Supplementary search techniques Reference lists of included studies Experts in the area

Citations searching
Author searching

Supplementary search techniques

Reviews for clinical studies routinely include additional search techniques such as checking reference lists of included studies and contacting experts in the area. Subsequent relevant studies may also be identified by a) searching citations of key relevant articles and b) searching for relevant publications of key authors in the area. These supplementary searches can be beneficial as demonstrated in the osteoporosis case-study. Of the 28 studies included in the review, 13 were identified through electronic searches, 8 were identified through references lists of the initial 13 studies, 4 were identified through contact with experts, and the final 2 were identified from a previous systematic review.

3.3 Using a sensitive or precise search strategy

When identifying publications describing clinical effects, it is important to maximise the sensitivity of the search. Consequently a range of synonyms for every concept are used in the search terms and an extensive search strategy is used as standard. The primary disadvantage of a sensitive search strategy is the potentially large number of results and the associated resources required. This is particularly relevant for HSUV searches. For example in the case-study, the use of the term 'quality of life' increased the number of non-relevant articles (i.e. produced a large set of results of lower precision). In addition, as discussed earlier, the nonspecific thesauri terms for HSUVs means that an extensive search strategy may not ensure that all relevant studies are identified, even when indexed on MEDLINE, as demonstrated in the case-study.

Taking a more pragmatic approach by using free-text terms for the relevant HRQoL measures and the omission of nonspecific MESH thesauri terms may be considered acceptable when used in conjunction with the supplementary search techniques (i.e. citation and reference searching etc.).

The full search strategy used to identify HSUVs for the case-study is provided in the NICE TSD 9 [4,5], and associated Supplemental Materials are available at: www.nicedsu.org.uk/TSD9%20HSUV%20values_FINAL.pdf
http://dx.doi.org/doi:10.1016/j.jval.2013.02.017

4. Inclusion/exclusion criteria when selecting studies for inclusion in the review

As in any review, eligibility criteria are defined prior to sifting results. Results are screened for inclusion against eligibility criteria using the standard stages (title, abstract, and full-text) and articles can be excluded if they do not meet one or more of the inclusion criteria at each sifting stage. However, with HSUV studies, there is an increased risk of excluding relevant studies at the title/abstract stage due to the non-standard reporting standards. For example, in the case-study, 8 of the studies identified by reference list checks had been excluded at the initial study selection stage (prior to full-text retrieval). In addition **Table 3** provides the inclusion criteria used in the case-study.

The downside of using non-exacting inclusion criteria is the increase in the number of full-text articles reviewed, which is compounded when iterative searches are performed to satisfy any revised requirements for the updated decision model.

Table 3 Inclusion criteria used in the case-study in osteoporosis

Inclusion criteria
Adults > 17 years of age
Men and postmenopausal women suffering from primary or secondary osteoporosis
Empirically estimated HSUVs using a recognised valuation technique (SG, TTO or VAS)
English language or translation

Key: SG: standard gamble; TTO: time trade-off; VAS: visual analogue scale

Papers which report utilities from other studies/papers are not usually included, but used to help identify the original source. Care should be taken to highlight separate papers that draw upon the same dataset or study.

There may be several sets of HSUVs required depending on the definitions of health states within the decision model (e.g. subgrouped by position of fracture in the case-study), and thus the decision process could involve multiple sets of potential references.

The appropriate inclusion criteria may be derived iteratively, at the scoping stage, depending upon the availability of studies containing relevant HSUVs. This is particularly pertinent to inclusion criteria relating to study quality. Whilst study quality will form an important consideration in ultimate study

selection decisions, it is difficult to state *a priori* strict criteria relating to methodological quality. For example, where there is little information at all reviewers may wish to include studies based on valuation of vignettes, or even clinical judgement, acknowledging and possibly exploring through sensitivity analysis the uncertainty that this type of evidence brings with it..

4.1 Preliminary data extraction

To assist in the study selection process and the refinement of the inclusion and exclusion criteria, it is recommended to undertake a preliminary data extraction of the following three key details:

1) The definition of the health of the population (e.g. age, sex, definition and severity of condition) and the individuals who complete the questionnaire (e.g. patient or carer). The former should reflect the health state definitions of the decision model, the latter should satisfy any reimbursement agencies requirements.
2) Details of the approach used to describe the health states, e.g. GPBM (such as EQ-5D, SF-6D, HUI); vignettes or scenarios; or direct measurement by time trade-off (TTO), standard gamble (SG) or visual analogue scale (VAS) including whether self-completed or proxy values are used.
3) Valuation method including who valued the health states (e.g. general public) and how (e.g. the elicitation technique such as TTO, SG or VAS).

5. Quality assessment and relevance of included studies

High quality clinical reviews generally include some form of assessment of the studies included and whilst it is equally important to review the quality of studies for HSUV reviews, there are no agreed reporting standards that encompass all the types of studies that could include HSUVs. The quality of studies included in these types of reviews can be difficult to assess and selecting a checklist based on the design of the primary study may be inappropriate. For example HSUVs may be secondary or tertiary outcomes in an RCT powered and designed according to a different primary outcome measure.

The ISPOR task force on estimating health-state utilities for economic models defined high quality HSUV in this context as “estimates that are aligned with the definitions of the economic model health states, are free from known sources of bias, and were measured using a validated method appropriate to the condition and population of interest and the perspective of the decision maker for whom the economic model is being developed” [14].

These four main considerations are explored in more detail in Table 4 which provides criteria for assessing the relevance and quality of HSUVs. Ultimately, any quality assessment is reliant on the quality of reporting standards in the primary studies included in the review, and if there is insufficient detail to assess the quality of the evidence, this should be noted as a shortcoming.

Table 4 Quality assessment and relevance checklist for studies including HSUVs

Criteria	Consideration
Relevance to the decision problem	
Population characteristics	How closely do the patient characteristics (e.g. age, sex, comorbidities, diagnosis, severity of condition) in the study match those modelled and those described in the decision problem?
Respondent selection and recruitment	Does this result in a population comparable to that being modelled?
Inclusion/exclusion criteria	Do these exclude any individuals? (e.g. the very elderly >80 years old are often not included in studies) How closely do the inclusion criteria match people who would receive the intervention in routine practice?
Quality assessment - free from bias	
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 the decision model
Response rates to the measure used	Are response rates reported and if so are the rates likely to be a threat to the validity of the estimated HSUVs for the health states?
Loss to follow-up	How large is the loss to follow-up (e.g. 1 year after fracture) and are reasons given? Are these likely to threaten the validity of the estimates?
Missing data	What are the levels of missing data and how are they dealt with? Are there details on the causes of the missing data? Again, could this threaten the validity of the estimates?
Utility values are measured and valued appropriately	
Appropriate use of valuation method	If valuation methods are used (TTO, SG, DCE, VAS) they are used appropriately? Does the valuation method provide preference based values anchored at 1 as equivalent to full health and 0 as equivalent to dead? Are adequate details of the valuation method provided to allow judgement on appropriateness?
Appropriate use of GPMB	Are adequate details of the PBM method provided (e.g. details given on the version used, the social tariff applied etc.)

	Was the GPBM delivered as intended? (e.g. wording and response options not changed) Is the measure used for the group it was intended (e.g. is an adult GPBM being used for children? Is EQ-5D-Y used with the adult tariff? Is a CS-PRM being applied to the intended group?)
Appropriate health-state description	If a health state is valued using a vignette can the accuracy of the vignette can be established, e.g. the process by which it was derived is described
In line with reimbursement agency requirements	
Utilities representative of the area of remit for the agency	Is the geographical area of recruitment relevant for the reimbursement agency (e.g. country patients recruited from)?
Measure used to describe the HSUVs	Does the measure used to collect the HSUVs match the requirements of the decision problem and reimbursement agency?
Population the HSUVs collected from	Who completes the measure (e.g. patient, proxy, judgement by clinician) and does it satisfy the requirements of the decision problem and reimbursement agency? Mode of administration – was it standard across participants? Was it in line with reimbursement agency requirements?
Population used to value the health states within the measure	Who values the health states (patients, carer, general population) and does this satisfy the requirements of the reimbursement authority of interest?
Technique used to value the health state	What technique is used to value the health stated (e.g. TTO, SG) and does this satisfy the requirements of the reimbursement authority of interest?

Source: adapted from Papaioannou et al.[5], page 694, with permission from Elsevier

6. Data extraction

While earlier a 'preliminary' data extraction was recommended to aid in the study selection process, once studies are identified as potential candidates for the review, a 'full' data extraction stage can be undertaken to assist in: informing the inclusion/exclusion of specific studies from the list of candidate studies; identifying if it may be possible to synthesise HSUVs and associated factors requiring consideration (e.g. heterogeneity); identifying if the data may require 'modifying' to match the exact needs of the health states within the decision model. The data extracted in the 'full' data extraction process may require replication for each of the individual health states within the model. For example, one individual study may provide HSUVs for both the initial effect and the effect one year post fracture for numerous fracture sites.

As for clinical effectiveness reviews, the data extraction forms are generally review specific, and it is recommended to pilot the design to determine the exact data to be extracted. There are however, many variables which are generic such as author, country of publication, study characteristics (inclusion/exclusion criteria), sample size, population characteristics (age, sex, condition and severity), and study setting. Information relating to the outcome of interest may differ depending on the measure. While ideally data extraction is undertaken by two independent reviewers and cross-checked (with dispute resolved by a third reviewer), in reality, the process is generally undertaken by a single reviewer and checked by a second reviewer.

An example data extraction form is provided (**Figure 4**) and this should be modified to match the requirements of the individual review (for example the disease-specific requirements will change for other conditions and models).

Figure 4 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	Participant characteristics used in study Age Sex Disease severity Any other relevant characteristics (e,g, the time since clinical diagnosis/onset of condition)

Method of elicitation of HSUVs – how and who e.g. vignettes/ health state descriptive system (EQ-5D), direct measurement Patient, clinician, general population	Valuation technique e.g. SG, TTO, VAS, and variant of the technique (e.g. Was a prop board used? Was top down titration used? What was the bottom anchor for the VAS?)

Respondent selection and recruitment	Response rates	Reasons for lost to follow-up	Any other potential problems with the study

HSUV descriptive statistics: per subgroup			
Sample size/ number of respondents	Mean (SD)	Median	Range

Source: Papaioannou et al. [4]

7. When and how to synthesise evidence

Methodological research exploring alternative way to synthesise HSUVs is in its infancy and to our knowledge, this is not a requirement of any reimbursement agency. However, if more than one set of the required HSUVs are available, and these are sufficiently homogenous (collected from similar populations, using the same measure and preference weights) then some form of pooling could be considered as a useful way of improving the precision of the estimates used.

Meta-analysis provides a means to pool data collected across a number of studies and produce a weighted average of the measure of interest (for example using the inverse of the variance, or sample size), thereby generating a more precise measure. There are two main methods:

- Fixed effects: This assumes that each study is estimating exactly the same value (the one true HSUV for the particular group of interest) hence any differences between studies should be explained by within study sampling error.

- Random effects: This assumes that each study is estimating a slightly different value (the true HSUV varies between studies), but the values follow a distribution across studies – with the studies providing a random sample from the distribution of true HSUVs. Any differences between studies should be explained by both within study sampling error and between study variation in the true value. The combined value therefore represents the mean of the possible true HSUVs.

Meta-analysis of HSUVs can be performed where a number of studies provide values on the same GPBM (such as meta-analysis on EQ-5D values (e.g. Peasgood et al, 2009 for osteoporosis states [2]; Doth et al, 2010 for pain states [15]), where the population is sufficiently similar (for example, they use levels of severity of a condition (e.g. mild depression) in a consistent manner across studies), and HSUVs are derived from the same preference weights (i.e. not using social tariff from different countries).

Meta-analysis is problematic where there is too much heterogeneity between values, as likely when considering all possible HSUVs for a particular condition. Whilst there are standard tests for heterogeneity within meta-analysis, judgement is also required as to whether it is conceptually appropriate to consider combining values generated using different methodologies, and based on different populations.

Meta-regression is a technique to explore possible causes of heterogeneity. The HSUVs are treated as the dependent variable with possible causes of heterogeneity explored through covariates, with values weighted by the precision of their estimate. However, modelling HSUVs is complex, and needs to address a number of issues:

- GPBM are a composite of a descriptive system and a social tariff
- HSUVs use a bounded scale, and cannot go above 1 (nor below the lowest tariff score for GPBM)
- Most HSUV studies report more than one mean HSUV (e.g. patients may complete more than one GPBM and they may complete the same GPBM at different times); consequently any meta-analysis of HSUVs needs to address the fact that these values will be correlated.

Furthermore, it is unlikely that the search will have identified sufficient mean HSUVs to enable the full exploration of possible group differences (which may additionally be inadequately modelled by a simple dummy variable). Meta-regressions with only a few studies and considerable study

heterogeneity run the risk of showing false positives [16], hence may be misleading. Whilst there are no hard and fast rules for the appropriate sample size in meta-regression, a ratio of at least ten studies to each covariate is often recommended [17].

Meta-regression may enhance understanding of heterogeneity within identified utility values, but should be treated with caution. Stricter inclusion criteria (although justified) for meta-regression or meta-analysis of HSUV, may help dealing with heterogeneity [18].

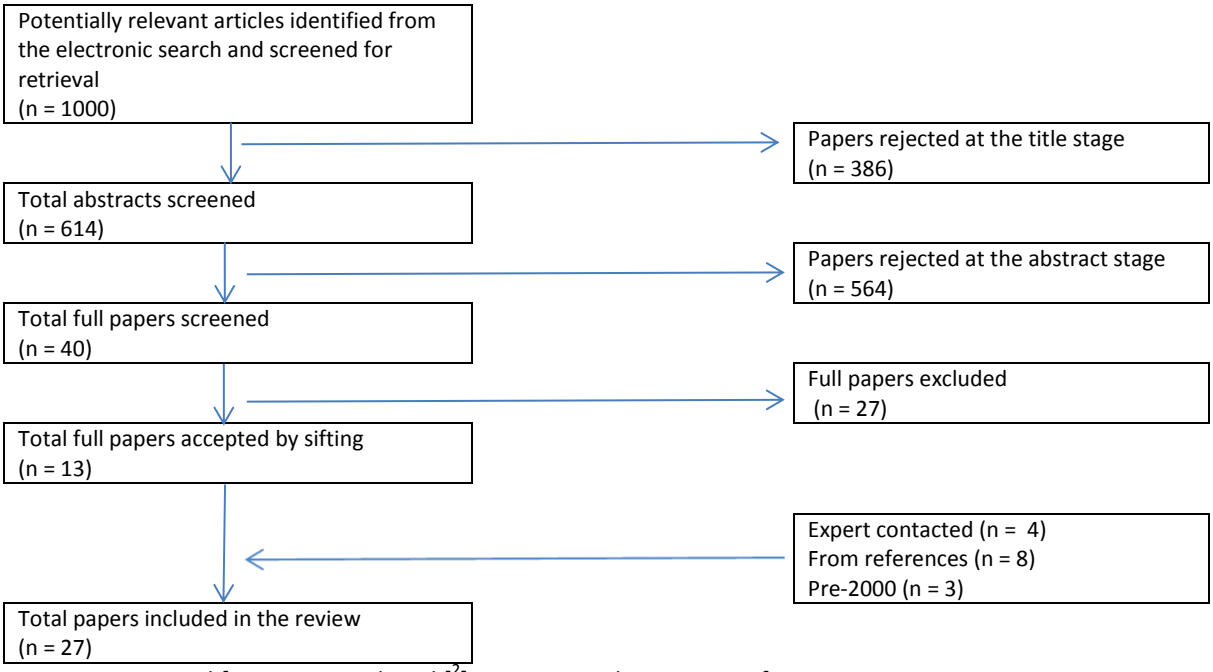
8. Presentation of results

As mentioned earlier, the search strategies and the inclusion / exclusion criteria used in literature reviews for HSUVs used in decision models can be an iterative process as the model evolves and the initial results are reviewed. Consequently, it is extremely important to document the process used throughout.

8.1 Searches/inclusion and exclusion of articles

The full search strategy and corresponding results for each of the iterative stages should be provided including the exact terms used in the searches, the databases searches, the numbers of hits, the number of potential relevant articles obtained from the supplementary searches, and the rational for exclusion of individual articles. Much of this information can be provided as appendices or as supplementary information depending on the reporting format or requirements of specific reporting standards. **Figure 5** is the flow diagram used to illustrate the study selection and exclusions for the osteoporosis case-study, which should be reported together with a tabulation of exclusion rational applied to the individual studies at each stage.

Figure 5 Osteoporosis case study: Statement flow diagram of study selection and exclusion



Source: Reprinted from Peasgood et al.[2], Figure 1, with permission from Springer.

8.2 HSUVs used in the decision model

Once the final selection has been made, it is good practice to provide tabulated mean HSUVs (and variance) used for each health state within the decision model, together with a full justification for its use. Any alternative values used in sensitivity analyses should be reported (together with the rationale for the value chosen e.g. the use of extreme values identified in the literature review). There are occasions when HSUVs have to be 'modified' in some way to match the definitions of the health states. Examples include: when combining HSUVs obtained from individuals with a single condition to represent the HSUV for a comorbidity, adjusting for age or an adverse event, or when increasing or decreasing the HSUVs to take account of the time since the acute event. The methodology and the rationale for the choice of values used should be reported explicitly and a range of sensitivity analyses should be conducted to reflect the uncertainty associated with the modified values.

8.3 Quality and relevance assessment

The checklists (e.g. CASP) completed during the reviewing process should also be available.

9. Summary

Literature reviews should be conducted early in the development process of the economic model to inform the possible impact and magnitude of effect on a particular health condition, and identify

appropriate preference-based measures. Reviews of health related quality of life literature can be resource intensive and scoping searches may be a useful technique to assess the sensitivity and precision of the search terms. Searching is an iterative process informed by the health states within the decision analytic model which may develop over time. Thorough and robust reporting is required at all stages to justify the final HSUVs selected.

Disclosure statement

This article is published in a special edition journal supplement wholly funded by Takeda Pharmaceutical International AG, Zurich, Switzerland.

Acknowledgement

The authors would like to thank Prof Jon Karnon, PhD of The University of Adelaide and Dr Andrew Lloyd, Phd of Bladen Associates Ltd for their editorial review.

Author contributions

RA reviewed the literature, and wrote the first draft and made the final edits to the manuscript. JEB made significant edits to the first and final draft of the manuscript. TP contributed to the initial draft of the manuscript and made edits to the final draft. SP made comments on the final draft of the manuscript.

Compliance with Ethical Standards

Funding This study was funded by an unrestricted grant from Takeda Pharmeceuticals International AG.

Conflict of interest Roberta Ara has no conflicts of interest. John Brazier has no conflicts of interest. Tessa Peasgood has no conflicts of interest. Suzy Paisley has no conflicts of interest.

REFERENCES

[1] Brazier J, Ara R, Chevrou-Severac H. W3: Utilities in economic evaluation: using best practices where international guidelines provide insufficient detail. Available from: www.ispor.org/Event/ProgramList/2016Vienna?type=Workshop accessed 13th April, 2017.

[2] Peasgood T, Herrmann K, Kanis JA, Brazier JE. An updated systematic review of Health State Utility Values for osteoporosis related conditions. Osteoporosis International. 2009 Jun 1;20(6):853.

[3] Kanis JA, Stevenson M, McCloskey EV, Davis S, Lloyd-Jones M. Glucocorticoid-induced osteoporosis: a systematic review and cost-utility analysis. Health Technology Assessment 2007; 11(7):iii-iv, ix-xi, 1-231.

[4] Papaioannou D, Brazier JE, Paisley S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. National Health Service, 2010. Available from http://www.nicedsu.org.uk.

[5] Papaioannou D, Brazier J, Paisley S. Systematic Searching and Selection of Health State Utility Values from the Literature. Value in Health 2013, 16(4):686-695.

[6] 01 Rowen D, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International regulations and recommendations. Current issue Pharmacoeconomics

[7] Guide to the methods of technology appraisal 2013 NICE. Available from: www.nice.org.uk/process/pmg9/chapter/foreword Accessed April 2017

[8] 02 Brazier J, Ara R, Rowen D, Chevrou-Severac H. A review of Generic preference-based measures. Current issue Pharmacoeconomics.

[9] 03 Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The role of condition-specific preference-based measures. Current issue Pharmacoeconomics.

[10] 05 Ara R, Rowen D, Mukuria C. The use of mapping to estimate health state utility values, Current issue Pharmacoeconomics.

[11] Critical Appraisal Skills Programme (2017). CASP (Systematic Review Checklist). Available at: http://www.casp-uk.net/

[12] Arber M, Garcia S, Veale T, Edwards M, Shaw A, Glanville J. (2016) Performance of search filters to identify health state utility studies. ISPOR 19th Annual European Congress. Vienna.

[13] Paisley S, Booth A, Mensinkai S. Health-Related Quality of Life Studies. Etext on Health Technology Assessment (HTA) Information Resources. Bethesda, MD: US National Library of Medicine (NLM), National Information Center on Health Services Research and Healthcare Technology (NICHSR), 2005.

[14] Wolowacz SE, Briggs A, Belozeroff V, Clarke P, Doward L, Goeree R, Lloyd A, Norman R. Estimating health-state utility for economic models in clinical studies: an ISPOR Good Research Practices Task Force Report. Value in Health. 2016 Oct 31;19(6):704-19.

[15] Doth AH, Hansson PT, Jensen MP, Taylor RS. The burden of neuropathic pain: a systematic review and meta-analysis of health utilities. Pain 2010; 149(2):338–344.

[16] Higgins J, Thompson S. Controlling the risk of spurious findings from meta-regression. Statistics in Medicine 2004, 23:1663–1682.

[17] Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to meta-analysis. New York: Wiley; 2011.

[18] Peasgood T, Brazier J. Is Meta-Analysis for Utility Values Appropriate Given the Potential Impact Different Elicitation Methods Have on Values? PharmacoEconomics 2015, 33(11):1101-1105.