Identification, Review and Use of Health State Utilities in Cost-Effectiveness Models:

Good Practices for Outcomes Research

ABSTRACT (205 words)

Cost-effectiveness models that present results in terms of a cost per quality-adjusted life-year (QALY) for health technologies are used to inform policy decisions in many parts of the world. Health state utilities (HSU) are required to calculate the QALYs. Even where clinical studies assessing the effectiveness of the health technologies collect HSUs to populate the cost-effectiveness model, which rarely happens, analysts generally need to identify at least some additional HSUs from alternative sources. When possible, these would be identified by a systematic review of the literature, but again this rarely happens.

In 2014, ISPOR established a Good Practices for Outcome Research Task Force for using HSUs in cost-effectiveness models. This task force report provides recommendations for researchers identifying, reviewing and synthesising HSUs for use in cost-effectiveness models, analysts that use the results in cost-effectivesss models, and reviewers that critically appraise the suitability and validity of the HSUs selected for these studies. The associated ISPOR Health State Utility Good Practices Task Force Minimum Reporting of Systematic Review of Utilities for Cost-Effectiveness (SpRUCE) Checklist provides criteria to judge the appropriateness of the HSUs selected for use and is suitable for use across different international settings.

Keywords: cost-effectiveness, systematic review, health state utility, health related quality of life.

Identification, Review and Use of Health State Utilities in Cost-Effectiveness Models: Good Practices for Outcomes Research

I. INTRODUCTION

Cost-effectiveness models that present results in terms of cost per quality adjusted life-year (QALY) for health technologies are used to inform policy decisions in many parts of the world. Health state utility (HSU) data are required to calculate QALYs. HSUs describe the value of a health state on a 0-1 scale, where one represents full health, zero represents states judged to be as bad as being dead and negative values represent states judged to be worse than dead. The preference values are usually obtained by elicitation techniques like standard gamble (SG) or time-trade-off (TTO) from a sample of the general population (though patient populations may be used).

If HSUs are not available from trials, and it is not feasible to conduct a study to collect this evidence, they are often sourced from the literature. This is problematic because analysts frequently cite dated evidence used in previous evaluations without undertaking basic quality checks of the data in the original source material, e.g., the relevance of the patient population, utility measure, elicitation method or sources of the preference weights. Furthermore, systematic reviews of the literature are rarely undertaken for HSUs, and current reporting standards are often poor (1).

Different samples, estimation methods and preference weights can result in different HSUs for the same health state (2,3,4). Selecting evidence in an ad hoc manner will result in unjustifiable

conclusions and raise the suspicion of 'cherry picking'. For consistency within a model, it is preferable that all health states are informed by evidence obtained using the same preference-based measure and preference weights, although this may not always be possible.

Where there are multiple appropriate HSUs for a particular health state or where it is not possible to identify all HSUs from the same measure, there may be a case for synthesising the data. Furthermore, it is likely that even the most appropriate HSUs may not exactly match the definitions of the health states within the model. Consequently, analysts frequently 'adjust' the data in some way to account for age, concurrent clinical events or adverse effects of treatment (5).

To address these issues, this report provides recommendations on the identification, critical appraisal and synthesis of HSUs from the literature, minimum reporting standards and the use of this evidence in cost-effectiveness models. It is the third ISPOR Health State Utility Good Practices Task Force Report. For detailed information on primary data collection and the derivation of mapping functions, please see the other reports (6,7).

II. SEARCH METHODS FOR IDENTIFYING HEALTH STATE UTILITIES IN THE LITERATURE

HSUs are available from a wide range of study designs, including randomised controlled trials, observational studies and economic evaluations (8). Systematic review guidance on how to search for studies systematically and transparently is useful for informing generic considerations, e.g., which databases to search and how to devise search strategies (9). However, searches for HSUs for models need to account for several requirements of the modelling process. These include: the iterative nature of model development, the scope of HSUs required, and judgments on the extensiveness of searches (10).

Iterative searching

Searches for HSUs are rarely a discrete, single activity at the outset of model development as final requirements may not be fully defined at that time. They are generally an iterative process involving multiple searches to identify the full scope of evidence required.

Initial scoping searches can inform early conceptualisations of the model and these early versions of the model will clarify specific information needs for further searches. For example, exploratory analysis may show that the model results are sensitive to certain HSUs, and insensitive to others. Future searches can then focus on the HSUs that influence the results. Consequently, interaction between the modeller and the information specialist is required to inform the evolving direction and scope of the iterative search activities.

Iterative searching can combine the more traditional in-depth search techniques with techniques to improve efficiency in order to explore a wide cross-section of potentially relevant evidence. Techniques to increase the efficiency of searching include initial, focussed searching to maximise the relevance of the search retrieval, e.g., by searching for relevant terms in the title only, followed by broader iterations of searching, e.g., by extending the search to abstracts. Guidance on iterative search techniques has been published by the NICE Decision Support Unit (11).

Scope of searches

The scope of evidence required should account for all health states and all aspects of treatment and management that might have an impact on HRQoL or might be affected by the intervention and comparators under consideration within the model. As such, multiple keyword search strategies may be required. For example, a cost-effectiveness model for the management of hypercholesterolaemia required HSUs for downstream events including stable and unstable angina, stroke and myocardial infarction (12).

A search approach that takes a systematic account of the full range of evidence requirements arising from the modelling framework is an important divergence from standard systematic review search methods. The latter are commonly associated with reviews of clinical effectiveness that focus on capturing evidence using a single search strategy defined by the population and intervention elements of the structured PICO (population, intervention, comparator, outcomes). A systematic account of the range of evidence to be retrieved should be determined by the requirements of the decision problem (See Figure 1). Factors to consider when identifying possible search criteria are provided in Table 1.

Table 1: Factors to consider when defining search criteria

Essential factors required to define search criteria
• Health state descriptions within the model
• Treatment effects of interventions and comparators of interest (including utility gains from treatment benefits and utility losses through adverse effects)
• Treatment effects and management at all stages of the clinical pathway included in the model
• Carer health state utilities
• Comorbidities
• Concurrent clinical events/sequelae
• General population norms
• Moderator variables that might affect quality of life, e.g., method of administration, treatment setting
Additional factors that may be relevant:
• 'Mapping functions' for estimating preference-based utilities from other HRQoL measures or clinical variables (See mapping section in this report.)

- The context within which the model will be used, e.g., geographical location or reimbursement agency criteria (23) [Rowen DL, Azzabi Zouraq I, Chevrou-Severac H & van Hout B (2017) International regulations and recommendations for utility data for health technology assessment. PharmacoEconomics, 35(Suppl 1), 11-19.]

Extent of searching

Exhaustive searching is a fundamental methodological precept of systematic reviews. While this is recommended practice for parameters of treatment effect (13), there is consensus that exhaustive searching for every model parameter is not an efficient use of resources (14,15).

To the extent that it is possible, it is important that the search process is 1) systematic and 2) explicitly stated to demonstrate that evidence has not been identified 'serendipitously, opportunistically or preferentially.' (11) Recommendations to achieve a minimum level of searching across all key model parameters, have been published (16) . These recommendations stress the need to undertake further searching if required or to provide justification if the minimum search level identified sufficient evidence.

Currently, there is no empirical definition of sufficient evidence or sufficient searching. In the absence of such definitions, the search objective should be 'to identify the breadth of information needs relevant to a model and sufficient information such that efforts to identify further evidence would add nothing to the analysis' (17). This concept is useful in informing heuristic judgments as to when to stop searching (18). Sufficiency checks include:

- Sensitivity analysis to understand the impact of HSUs on model outputs. Search activities can prioritise HSUs to which outputs are most sensitive.
- The availability of evidence. Extensive searching is not of value where there is minimal or a lack of appropriate evidence.

Search tools

Guidance on how to search for studies for inclusion in systematic reviews of HSUs provide details on how to search general biomedical databases, such as Medline and specialist databases (including the TUFTS Database, ScHARRHUD and the HERC Mapping Database (Health Economics Research Centre, Oxford, UK)) (19). The ISSG (InterTASC Information Specialist Subgroup) Search Filter Resource (20) provides HSU filters for use with databases such as Medline (21). Further guidance on searching is useful in adapting the search process for HSU systematic reviews to the specific requirements of cost-effectiveness models (11,16).

REVIEWING PROCESS OF HEALTH STATE UTILITIES

After completing the iterative literature searches and identifying articles that satisfy the inclusion criteria, some general recommendations can be considered as a starting point to review the articles. Initial considerations include the quality and appropriateness of the data. Depending on the target reimbursement agency or audience, additional considerations may include the choice of preference-based measure and/or source of preference weights, the study setting, and whether to allow evidence from another measure (Figure 1).

Quality of the data

Studies should be reviewed for evidence of methodological flaws or bias and limitations using the following as a minimum:

a. Precision of the evidence: The precision of the data will be reflected in the variance, which is related to sample size.
b. Response rate: The generalisability and validity of the evidence may be compromised if a substantial proportion of eligible subjects declined to participate.
c. Loss to follow-up and missing data: The rates of losses to follow-up may compromise the representativeness of the final sample. The levels of missing data, whether these can be

considered as missing at random, and how researchers deal with these, must be reported. This is particularly important in longitudinal evidence where data are assessed at interim points over time.

Appropriateness of the data

Patient characteristics/health state definitions

First and foremost, the population of the study must be comparable to the population modeled. Model's health states are often defined in terms of objective clinical measures. It may be necessary to have HSUs for health states defined by stage or severity of disease, comorbidities, age, gender, ethnicity, adverse events, or complications and sequelae. In chronic conditions characterised by symptom exacerbations, e.g., Crohn's disease or gout; or multiple discrete events, e.g., a transient ischemic attack or asthma, HSUs can fluctuate over time. Thus, it is important to consider timing of data collection, e.g., how close in time was the event and data collection point, and is this likely to result in statistically different HSUs? The use of any medications that are likely to have an independent effect on HSUs (either detrimental or beneficial) should be considered.

Preference-based measure and elicitation method

It is common practice for HSUs to be based on patients completing an HRQoL measure with the general public providing weights for the measure using techniques, such as TTO or SG (22). In general, when using HTA for decision-making on reimbursement or new technologies, societal weights are preferred over patients' (23). There are deviations from this and some decision makers prefer the weights from patients rather than the public e.g., the Dental and Pharmaceutical Benefits Agency in Sweden. Sometimes it is necessary to use proxy assessment, sometimes condition-specific measures are preferred to generic, and some agencies prefer all HSUs from the same measure, e.g., NICE and the Dutch National Health Care Institute (23).

An important aspect is the extent to which the measure is valid or appropriate for the condition. A measure should be sensitive to changes in the domains of health likely to be affected by the condition. For example, when evaluating interventions for mental conditions affecting self-esteem or social relationships, it is important that the measure can capture changes in these (24). In some conditions, certain measures have been shown to be insensitive, e.g., EQ-5D in hearing loss and some visual disorders (25).

Consistency of evidence

Using a single measure (and preference-weights) for all HSUs within a model removes variance due to different valuation methods, populations, etc. However, it is not always possible to identify all HSUs from a common measure. There may be a trade-off between the desirable characteristics of the HSUs or a specific measure, and coverage of the most important health states in the model in terms of effect on the incremental cost-effectiveness ratio (ICER).

The final evidence used may be selected by trading off 'ideal' characteristics (see Figure 1) that are likely to differ across models. Where all HSUs needed are not available from a common source, consistency of the measure is a priority, subject to the robustness of the data. In some instances, the exact patient characteristics and timing of the data collection may outweigh the specific measure. The final selection should be transparently reported and justified by the use of pre-determined criteria. Any suitable alternatives should be considered in sensitivity analysis.

III. SYNTHESIS OF HEALTH STATE UTILITIES

Why undertake a synthesis?

There are often multiple published HSUs for a given health state. To use one source per state is not making best use of the available evidence. The aim of synthesis is to generate a more accurate

estimate of the mean HSU and the associated uncertainty, and to improve the generalisability of findings.

When should a synthesis be undertaken?

Before undertaking a formal synthesis of HSUs, it is important to consider whether there are enough HSUs and whether the studies are sufficiently homogenous for the aggregation to be meaningful. For meta-regression of effectiveness, a minimum of four studies in a categorical subgroup variable has been suggested (26), while more are required to conduct significance testing.

Heterogeneity can be a major problem. Peasgood et al. (2015) identified considerable variability in HSUs arising from differences between: measures (EQ-5D vs. SF-6D), valuation method, (TTO vs. SG), the types of anchors used, the country of the valuation, and who provided the preference weights (patient vs. general population) (27). The large number of sources of variation can imply that any formal synthesis is not meaningful particularly if they exceed the number of HSUs used.

What's the role of mapping **in evidence synthesis?**

Mapping can expand the number of relevant HSUs available for synthesis in two situations. The first is studies using health or HRQoL measures that do not generate preference-based HSUs. The second is where HSUs are obtained using different preference-based measures, or different valuation techniques.

In both cases, there may be functions that map or cross-walk from one measure onto a generic preference-based measure, e.g., EQ-5D (28). However, mapping functions increase uncertainty and can produce systematic error in estimation (6). Mapping functions are generally used in situations where patient level data is available, although it is possible to map from mean HSUs.

Methods of synthesis

Syntheses aim to estimate the absolute or relative impact of each health state on the corresponding HSU. The methods for synthesis are at an early stage of evolution. Therefore, we are limited in the recommendations that we can currently make. There are two broad approaches. One involves applying strict eligibility criteria to studies included in the analysis in order to reduce heterogeneity, such as limiting HSUs to those obtained from the same measure and specific sub-groups, e.g., mild, moderate or severe depression.

This is appropriate where there are sufficient numbers of HSUs meeting the criteria. For example, Peasgood, et al. (2015) excluded all non EQ-5D evidence (to meet NICE's preferred measure) and combined nine studies to estimate mean HSUs (27). Considerable unexplained heterogeneity remained despite using the same measure, which raised concerns about the relevance of the estimates for use in cost-effectiveness models.

When there are not enough studies using the same method on a sufficiently homogenous population, more sophisticated methods are needed. The second approach attempts to explicitly model the impact of heterogeneity on the HSUs using meta-regression. For example, Bremner and colleagues (2007) estimated a linear mixed-effects model in prostate cancer to estimate coefficients for disease stage, symptoms, severity, and valuation methods (29). The authors acknowledged problems with over-predicting HSUs at the lower end and predicting HSUs greater than one.

A study in colorectal cancer used a similar linear mixed logit model and compared it to a Bayesian logit model-based model. They found the latter gave a better fit, although the coefficients need transforming for use in cost-effectiveness (30). In both studies, considerable

heterogeneity remained, partly because the models were limited by the variables published in the studies, and partly because the authors did not have access to individual level data.

Meta-regression methods require a lot of data to control for the different sources of variation between studies. Ten studies per covariate has been suggested in the literature, but this may not be realistic for many indications (31). Methodological research is needed into methods of meta-regression when synthesising HSUs and when they are appropriate.

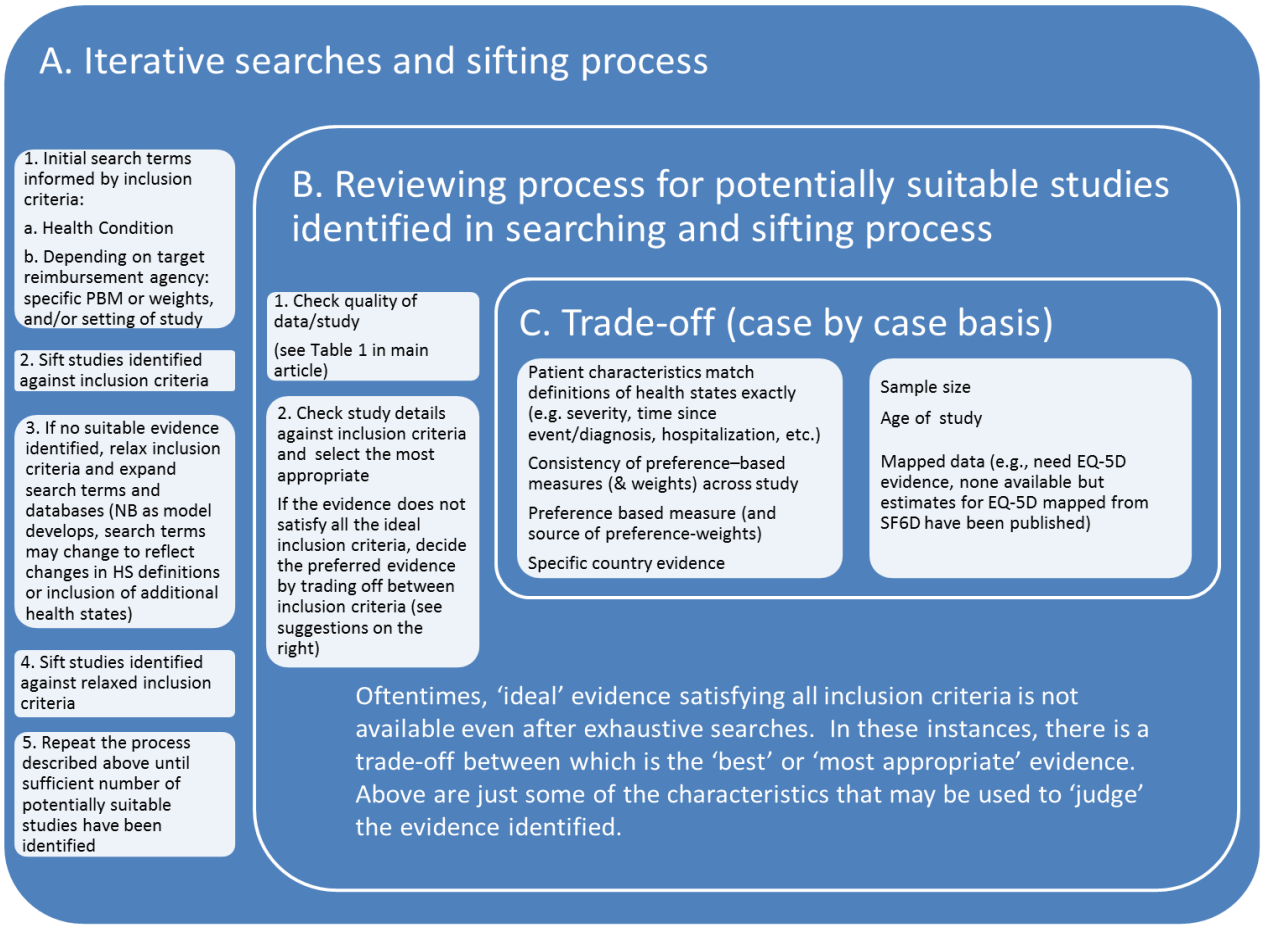
Another under-explored source of variation is that evidence differs by country of patient recruitment. This may be due to differences in patient characteristics that can be controlled for in meta-regression. It can also be due to differences in country specific preference-weights for measures like EQ-5D. However, oftentimes the preference-weights in one country are used in another, e.g., UK EQ-5D preference-weights are used in submissions to the US Institute for Clinical and Economic Review.

There is no standard way to re-weight from published values, and this can only be done with access to individual level data. In addition, there may be a country-specific effect from the general health of patients or the impact of the healthcare system more generally. The importance of these different sources of variation and how to deal with them needs to be further explored.

Current experience using formal synthesis methods is limited for HSUs. For pragmatic reasons, many of the more complex techniques commonly used in the clinical efficacy literature may have little role when synthesising HSUs due to the limited number of studies and the high degree of heterogeneity in the valuation methods and patient populations. However, with a growing literature, there will be increasing opportunities to use meta-analysis with HSUs.

IV. MINIMUM REPORTING STANDARDS FOR LITERATURE REVIEWS AND MODELLING REPORTS

We recognize the challenge of extensive documentation when multiple literature reviews are necessary and models encompass multiple conditions/co-morbidities. However, the fundamental tenets of systematic reviews, such as systematic search, critical appraisal and transparency of reporting, as described in the ISPOR CHEERS report, are still critical to the success of the review (32). The iterative nature of the search and review process is outlined in Figure 1.



In Table 2, we outline criteria to support minimum reporting standards for the systematic review of HSUs for cost-effectiveness - the ISPOR HSU Good Practices Task Force Minimum Reporting Standards of Systematic Review of Utilities for Cost-Effectiveness (SpRUCE) Checklist (ISPOR SpRUCE Checklist). These criteria were designed to help reviewers identify if HSU selection for the model was transparent and appropriate. The checklist has five sections that refer to the search

strategy, the review process, the data extracted from each study, the rationale for the final HSU, and their use in the model. While the ISPOR SpRUCE Checklist provides a minimum set of reporting standards for HSUs in models, a greater level of detail is likely needed to proceed to peer-reviewed publication of a systematic review (33).

Table 2: ISPOR HSU Good Practices Task Force Minimum Reporting Standards of Systematic Review of Utilities for Cost-Effectiveness (ISPOR SpRUCE Checklist)

Criteria	Description
Search Strategy	
Search terms and scope	The final search strategy should be adequately defined, and appropriate databases included in the search.
Study selection criteria	Explicit criteria for study identification/inclusion should be described and applied, such as patient group of interest, relevant age range and stage of disease/severity etc.
Review Process	
Quality check	Quality criteria for reviewing studies is explicitly stated and applied.
Assessment of relevance	Relevance of HSUs to model and target reimbursement agency described.
Data Extracted (Reporting of variables)	
Population/patient characteristics	Include relevant patient characteristics such as age, sex, comorbidities, diagnosis, severity of condition.[#]

Measure used to describe the HSUs	Provide the name of the actual measure.
Preference weights	State the technique used to value the health state e.g., TTO, SG, and the country. Provide the reference.
Descriptive statistics of HSUs	Include the mean and variance around any HSU used in the model.
Response rates to the measure used*	Report if response rates are likely to be a threat to validity.
Loss to follow-up/ missing data*	Report loss to follow-up, e.g., 1 year after fracture, and missing data especially if they may threaten the representativeness of the HSUs.
Original reference	The original source for the HSUs should be referenced (NOT a previous economic study that has used the evidence).
Selection/estimation of final health state utilities	
Basis for selecting HSUs	The rationale for selecting the HSUs used in the model should be justified.
Method used to combine estimates	Where HSUs are combined, the analytic methods should be described, e.g., meta-analysis.
Methods used when applying the health state utilities in model	
Actual HSUs used	Report all actual HSUs used in the model together with the associated measure.
Adjustments or assumptions	Clearly describe any adjustments or assumptions relating to the use of HSUs in the

	model. Report both the raw and final values used with worked examples, if required to clarify the method used to adjust the data.

*Extract and report if relevant.

[#] The original primary source should be checked rather than relying on the use of data from a similar economic model without checking relevance

SG: standard gamble; HSU: health state utility; TTO: time trade-off

Search strategies

Ideally, the search and selection methods used in a systematic review should be described in a protocol prior to study initiation. While initial searches may be somewhat cursory, HSUs found to be important, e.g., through sensitivity analysis, should drive a more comprehensive search strategy. Associated reports should specify the terms used in search strategies and the databases reviewed. Additional non-standard search strategies, e.g., hand searches, non-peer reviewed literature or HTA type submissions should be described.

Review Process

The process for screening and determining the eligibility of studies should be reported together with the number of reviewers involved and how disputes were resolved. The iterative search process and associated inclusion criteria could be summarised in a PRISMA flow diagram (34). Studies reporting HSUs that met the inclusion criteria, yet were not selected after the 'trade-off' process, should be listed.

Data extraction

Upon identification of studies that meet the inclusion criteria, data used to assess the appropriateness of the HSUs should be extracted and summarized (See Figure 1, Section 2). For

modelling reports, data extraction could be limited to those studies that were included as a model input.

Elements of the studies providing the HSUs used should be described in the review, such as study design, e.g. observational or clinical trial, whether it was a clinical study versus patient survey, and possible study limitations, such as aspects of the design that may promote placebo effects that can inflate HSUs. It is important to identify and reference the original source document. Secondary references are a common issue in reviews of HSUs and modelling reports. If the authors of a study have been contacted for clarification or even original data, this should be documented.

Selection/estimation of final HSU for model

If a review results in the identification of multiple appropriate HSUs, the following should be reported: 1) selection - justify the rationale for the selection of the best evidence or 2) estimation - the methods for combining the evidence should be reported and justified, e.g. meta-analysis, with tests for heterogeneity that support combining the data.

V. USING HEALTH STATE UTILITIES IN COST-EFFECTIVENESS MODELS

This section describes issues related to the use of HSUs within cost-effectiveness models and associated recommendations (See Table 3).

Discrete health states or discrete event simulation?

In a cost-effectiveness model, HSUs are most commonly assigned to a set of discrete health states using state transition models. However, the guidance here extends beyond these. Modelling techniques, such as discrete event simulation (DES), can represent the utility effects of all changes in clinical status through the estimation of HSUs as a function of clinical status. When

conceptualising a model structure, the number of discrete health states required to capture changes in clinical status that result in important changes in HSUs should be carefully considered (35).

There is no consensus on defining important changes in HSUs. These will likely vary by condition and utility measure, but the basis on which important changes in HSUs inform the model structure should be stated explicitly. The required number of discrete health states may lead to a decision to use DES (36). If a simpler model structure is implemented that does not represent all potentially important HSUs, the expected effects of such omissions should be examined and discussed.

Individual or function-based health state utilities?

HSUs may be estimated individually, analysing the data for each health state separately or a relevant function may be generated. If both options are available, the choice of data used in the base case model analysis should be informed by: 1) the relevance of the data (see Section 2), and 2) the reliability of the analyses, e.g., the precision of the mean HSUs and the validity of estimated functions. Again, the rationale for the final choice should be explicitly stated.

Comorbidity utility effects and the use of general population norms

HSUs should reflect HRQoL effects associated with the condition of interest, but also any comorbidities unrelated to that condition. The utility effects of comorbidities are real and should be represented in HSUs. The consequences of omitting these effects will be greatest in evaluations of interventions that increase life expectancy because QALY gains will be overestimated if the utility effects of unrelated comorbidities are not represented (37, 38).

It is reasonable to assume that mean HSUs represent comorbidity effects at the mean age of the utility study population. HSUs at younger and older ages should be adjusted to reflect age-related comorbidity utility effects. Age-specific HSUs should be estimated using the appropriate

385 'multiplier'. For example, if a condition specific HSU of 0.72 is derived from a study sample
386 with mean age 70 years, and the general population norm at age 70 years is 0.8, the multiplier is
387 0.72/0.8 = 0.9. Age-specific HSUs are then estimated for other ages using the multiplier, e.g., if
388 the general population norm at age 71 years is 0.79, the age-adjusted condition specific HSU at
389 age 71 years is 0.79 x 0.9 = 0.711. If the intervention is prophylactic and suitable data are
390 available, it is preferable to utilise condition-specific age-adjusted HSUs for the 'condition-free'
391 health state. Evidence suggests these may be lower than general population norms (39).
392

393 **Table 3: Methodological recommendations for using health state utilities in cost-**
394 **effectiveness models**

Issue	**Recommendation**
Individual or function-based HSUs?	Decisions should be informed by the relevance of 1) the data, e.g., the study population, 2) the utility measure and alignment with the model's health states, and 3) the reliability of the reported analyses, e.g., the precision of the mean HSUs and the validity of estimated utility functions.
Comorbidity utility effects and the use of general population norms	Mean HSUs represent comorbidity utility effects at the mean age of the utility study population. Age-specific comorbidity effects should be estimated using age-specific population norms. If the intervention is prophylactic and suitable data are available, it is preferable to utilise condition-specific age-adjusted HSUs for the 'condition-free' health state.
Treatment-related adverse events	The extent to which the utility effects of adverse events are captured by the data used to estimate a model's non-adverse event HSUs should be assessed. If adverse event HSUs are required, the range of HSUs to be estimated should be informed by their expected impact on cost-effectiveness.

Concurrent clinical events	The multiplicative method should be used to handle the utility effects of multiple concurrent clinical events.
Acute clinical event	In the absence of data collected around the event, plausible HSUs for the direct effects of acute events should be multiplied by the expected duration of the direct effects to assess the sensitivity of cost-effectiveness to these utility effects.
Sensitivity analysis	One-way and multi-way sensitivity analyses of HSUs should be undertaken. The difference method should be used to maintain appropriate ordering of HSUs in probabilistic sensitivity analyses.

Treatment-related adverse events and concurrent conditions

The need to estimate the disutility associated with adverse event reflects the extent to which this is already captured in the HSUs used for the model's health states. If individuals experiencing adverse events were less likely to return utility data, the disutilities of adverse events are likely to be underestimated. Alternatively, little data may be available on high impact, but uncommon, adverse events. In these cases, additional literature should be sought to estimate the disutility of adverse events, noting that the original HSUs may partially reflect adverse event effects.

The adverse event HSUs to be estimated should be justified with reference to incidence rates in the different treatment groups, their severity and duration, and the expected sensitivity of the cost-effectiveness results to the adverse event HSUs. A wider range of adverse event HSUs should be estimated as the expected impact of the HSUs on the cost-effectiveness results increases. The estimated effects should reflect the expected duration and pathway of the adverse events.

Individuals may also experience concurrent clinical events related to the condition of interest, e.g., diabetic patients may experience both cardiovascular disease and retinopathy. Approaches to handling the utility effects of multiple concurrent clinical events include: 1) subtracting the sum

of the estimated utility decrements for overlapping events from baseline HSU (additive); 2) multiplying the baseline HSU by the product of the ratios of the HSUs for individuals with and without the clinical events (multiplicative); and 3) the use of the lowest HSU across the clinical events (minimum) (38).

A review of 11 studies that used HSUs for single health conditions to estimate HSUs for concurrent health conditions found the minimum approach overestimated all observed HSUs and the multiplicative method was generally preferred to the additive method (38). The review noted the potential value of regression-based predictions of concurrent utility effects, whilst recognising the need for further research to validate regression approaches. On the basis of the existing evidence, the multiplicative method (using an appropriate multiplier) is the recommended approach.

Acute clinical event utility effects

Acute clinical events, such as asthma exacerbations and bone fractures, may be associated with large utility decrements due to high levels of pain or discomfort. However, it is rarely the case that respondents complete utility measures during the time period over which the effects of such acute events are experienced. The best solution to this problem is to tailor data collection to the events of interest (7). The impact of omitted utility effects is likely to be greatest for acute events that occur on a regular basis, such as asthma exacerbations and angina attacks. The timing of data collection for HSUs used from the literature should be assessed for appropriateness before used in the model.

To assess the sensitivity of the model outputs to acute event utility effects, analysts should generate plausible HSUs based on the expected clinical effects of the event, e.g., defining expected dimension of health levels from a multi-attribute utility instrument. The estimated HSU

can be multiplied by the expected duration of the effects to estimate the QALY loss per acute event, which can be applied to each occurrence of the event in the cost-effectiveness model.

Sensitivity analysis

Uncertainty around the mean HSUs (including population norms) should be represented by parametric probability distributions (40). Lower and upper confidence limits can be used in deterministic sensitivity analyses and random samples generated from the distributions for probabilistic sensitivity analyses (PsA). Uncertainty around HSUs should generally be represented by a standard beta distribution, which is bounded by 0 and 1. However, alternative lower and upper limits should be defined if a negative HSU is possible (41).

One-way sensitivity analyses should be undertaken to identify the HSUs to which the model results are most sensitive. Relevant multi-way sensitivity analyses include combined analyses of all HSUs, taking care to select combinations of lower and upper HSUs that move the cost-effectiveness results in the same direction, i.e., all selected HSUs either improve or worsen cost-effectiveness.

Ordered HSUs refer to pairs of HSUs in which there is an absolute belief that the true expected HSU for one state is higher than the true expected HSU for another state, e.g., a pre-diabetes HSU is higher than a diabetes HSU. In PsA, inconsistent HSUs can be sampled if there are overlapping probability distributions for ordered HSUs, e.g., a higher value could be sampled for the diabetes than for the pre-diabetes.

To avoid sampling inconsistent HSUs, the difference method should be used (42). This involves generating a probability distribution of the difference in the HSUs of two ordered parameters. In PsA, one of the ordered parameters is sampled, and the difference between the two HSUs is then added to the sampled value to generate the second HSU.

CONCLUSION

This report provides good practice guidance when identifying, reviewing and synthesising HSUs from the literature and using HSUs in cost-effectiveness models. Historically, analysts have paid insufficient attention to this parameter, often simply taking evidence used in previous models or those from a known source with no transparency or justification of choice. While the time and resources available for populating cost-effectiveness models will always be limited, the HSUs can be just as important as other parameters in models.

It is not always feasible or necessary to undertake comprehensive literature searches, but it is essential to report the search methods and the criteria used to review studies transparently. The processes for searching and reviewing are iterative as the scope of a search will depend on the literature available. It may be necessary to relax the search terms and inclusion criteria to allow more measures to identify appropriate evidence.

Any review criteria should be stated a priori as there are often trade-off decisions between criteria. Where all HSUs needed are not available from a common source, consistency of the measure is a priority, subject to the robustness of the data. Finally, searching and reviewing should be undertaken as part of the model development, since the results can influence the structure of the model and the sensitivity of the ICER can inform prioritisation of searches to the HSUs that are influential.

Due to the increase in the evidence base reporting HSUs, there may be more than one relevant HSU for each health state. Analysts should consider meta-analyses to generate more representative estimates (as for any other model parameter) or meta-regression to utilise the full range of evidence from heterogeneous studies.

While the literature is growing, there are often 'gaps' in the evidence. Analysts frequently adjust HSUs to account for adverse events, comorbidities and age. Analysts should report any issues with the evidence sources, the methods used to adjust the data, and the actual HSUs used in the model explicitly and transparently to enable readers to review the implications of the decisions made. The uncertainty in the HSUs should be captured appropriately.

These good practice task force recommendations and the ISPOR SpRUCE Checklist offer a structured and transparent basis for identifying and reporting the HSUs used in a cost-effectiveness model.

References

1. Brazier J, Ara R, Chevrou-Severac H. ISPOR 19th Annual European Congress, 29 October – 2 November 2016. Austria Center Vienna, Austria. W3: Utilities in economic evaluation: using best practices where international guidelines provide insufficient detail.

2. Janssen B, Szende A. (2014) Population Norms for the EQ-5D. In: Szende A, Janssen B, Cabases J. (eds) Self-Reported Population Health: An International Perspective based on EQ-5D. Springer, Dordrecht

3. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health economics. 2004 Sep 1;13(9):873-84.

4. Richardson J, Iezzi A, Khan MA. Why do multi-attribute utility measure produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects, Quality Of Life Research: An International Journal Of Quality Of Life Aspects Of Treatment, Care And Rehabilitation 2015; 24: 2045-2053.

5. Tosh JC, Longworth LJ, George E. Utility values in National Institute for Health and Clinical Excellence (NICE) technology appraisals. Value in Health 2011 Feb 28;14(1):102-9.

6. Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to Estimate Health-State Utility from Non–Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. Value Health 2017; 20(1):18-27.

7. Wolowacz SE, Briggs A, Belozeroff V, et al. Estimating health-state utility for economic models in clinical studies: an ISPOR good research practices task force report. Value Health

2016;19(6):704-719. http://www.ispor.org/Estimating-Health-State-Utility-Economic-Models-Clinical-Studies-guidelines.asp

8. Paisley S. Classification of evidence in decision-analytic models of cost-effectiveness: A content analysis of published reports. International Journal of Technology Assessment in Health Care 2010: 26(4):1-5.

9. Higgins JPT, Green S. (2011) Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [Updated March 2011]. Version 5.1.0. 2011. Cochrane Collaboration. https://www.ispor.org/PEguidelines/index.asp (Accessed November 2017).

10. Paisley S. (2012). Identifying evidence to inform decision-analytic models of cost-effectiveness : a qualitative study of information seeking processes and behaviour. Ph.D. Thesis. Sheffield: University of Sheffield.

11. Kaltenthaler E, Tappenden P, Paisley S, Squires H. (2011). Identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models. NICE DSU Technical Support Document 13. Sheffield: NICE DSU. Available from http://www.nicedsu.org.uk (Accessed 27 October 2018).

12. Ara R, Tumur I, Pandor A et al. Ezetimibe for the treatment of hypercholesterolaemia: a systematic review and economic evaluation. Health Technology Assessment 2008;12(21). doi:10.3310/hta12210.

13. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision-analytic modeling in health-care evaluation : report of the ISPOR Task Force on Good Research Practices - Modeling Studies. Value in Health 2003; 6(1):9-17.

14. Cooper NJ, Sutton AJ, Ades AE, Paisley S, Jones DR. Use of evidence in economic decision models: practical issues and methodological challenges. Health Economics 2007;16(12):1277-1286. doi:10.1002/hec.1297

15. Golder S, Glanville J, Ginnelly L. Populating decision-analytic models: The feasibility and efficiency of database searching for individual parameters. International Journal of Technology Assessment in Health Care 2005; 21(3):305-311.

16. Paisley S. Identification of Evidence for Key Parameters in Decision-Analytic Models of Cost Effectiveness: A Description of Sources and a Recommended Minimum Search Requirement. PharmacoEconomics 2026; 34(6): 597-608. doi:10.1007/s40273-015-0372-x

17. Glanville J, Paisley S. (2010). Searching for evidence for cost-effectiveness decisions. In I. Shemilt, M. Mugford, L. Vale, C. Donaldson, & K. Marsh (Eds.), Evidence-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice (2nd ed., pp. 79-92). Wiley-Blackwell.

18. Fenwick E, Knies S, Rothery C, Koffijberg E. Value of Information analysis for research decisions: Emerging good practice recommendations from the ISPOR VOI task force ISPOR 20th Annual European Congress, Scottish Event Campus, Glasgow, Scotland. Available from: https://www.ispor.org/docs/default-source/task-forces/value-of-information-workshop-glasgow-2017.pdf?sfvrsn=88157374_2. Accessed 7th October 2018.

19. Papaioannou D, Brazier J, Paisley S. (2010). The identification, review and synthesis of health state utility values from the literature. NICE DSU Technical Support Document : 9. Sheffield:

582 NICE DSU.Available from: nicedsu.org.uk/technical-support-documents/ (Accessed 7 October
583 2018).
584
585 20. InterTASC Information Specialists' Sub-Group. (2010) ISSG Search filter resource.
586 https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home
587
588 21. ArberM, Garcia S, Veale T, et al. Performance of ovid medline search filters to identify health
589 state utility studies. International Journal of Technology Assessment in Health Care
590 2017; 33(4);472-480. doi:10.1017/S0266462317000897
591
592 22. Brazier JE, Longworth L. NICE DSU Technical Support Document 8: An introduction to
593 the measurement and valuation of health for NICE submissions. 2011. Available from
594 http://www.nicedsu.org.uk (Accessed 25th April 2018)
595
596 23. Rowen D, Zouraq IA, Chevrou-Severac H, van Hout B. International regulations and
597 recommendations for utility data for health technology assessment. Pharmacoeconomics. 2017
598 Dec 1;35(1):11-9.
599
600 24. Brazier JE, Connell J, Papaioannou D, et al. Validating generic preference-based measures of
601 health in mental health populations and estimating mapping functions for widely used specific
602 measures. Health Technology Assessment 2014, volume 18(34).
603
604 25. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of
605 health-related quality of life in NICE decision-making: systematic review, statistical modelling
606 and survey. Health Technol Assess. 2014 Feb;18(9):1-224. doi: 10.3310/hta18090.
607

26. Fu R, Gartlehner G, Grant M, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011;64(11):1187–97.

27. Peasgood T, Brazier J. Is meta-analysis for utility values appropriate given the potential impact different elicitation methods have on values? Pharmacoeconomics 2015; 33: 1101-1105.

28. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European Journal of Health Economics 2010;11(2): 215-225.

29. Bremner KE, Chong CA, Tomlinson G, Alibhai SM, Krahn MD. A review and meta-analysis of prostate cancer utilities. Med Decis Making 2007;27(3):288–98.

30. Djalalov S, Rabeneck L, Tomlinson G, et al. A review and meta-analysis of colorectal cancer utilities. Med Decis Mak. 2014;34(6):809–18.

31. The Cochrane Collaboration: The Cochrane Manual Issue 4, 2008. http://www.cochrane.org/admin/manual.htm

32. Husereau D, Drummond M, Petrou S, et al. Consolidated health economic evaluation reporting standards (CHEERS)—Explanation and elaboration: A report of the ISPOR health economic evaluations publication guidelines good reporting practices task force. Value in Health 2013;16:231-50.

33. Santos J, Palumbo F, Molsen-David E et al. ISPOR Code of Ethics 2017 (4th Edition) Value in Health 2017; 20:1227-1242.

34. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of clinical epidemiology 2009; 62: e1-34.

35. Roberts M, Russell LB, Paltiel AD, et al. Conceptualizing a model: A report of the ISPOR-SMDM modeling good research practices task force-2. Value Health 2012;15:804-11.

36. Karnon J, Afzali HH. When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. Pharmacoeconomics 2014 Jun 1;32(6):547-58.

37. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value in Health 2010 Jul 1;13(5):509-18.

38. Ara R, Wailoo AJ. Estimating health state utility values for joint health conditions: a conceptual review and critique of the current evidence. Medical Decision Making 2013 Feb;33(2):139-53.

39. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value in Health. 2011 Jun 1;14(4):539-45.

40. Briggs AH, Weinstein MC, Fenwick E, et al. Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM modeling good research practices task force-6. Value Health 2012;15:835-42.

662 41. Filliben J. Exploratory Data Analysis. In: Croarkin C, Tobias P, editors. Engineering Statistics
663 Handbook. NIST/SEMATECH, 2012.
664 https://www.itl.nist.gov/div898/handbook/eda/section3/eda366h.htm
665
666 42. Ren S, Minton J, Whyte S, Latimer NR, Stevenson M. A new approach for sampling ordered
667 parameters in probabilistic sensitivity analysis. PharmacoEconomics 2017 Oct;28:1-7.
668