
Article:
Ara, R.M. orcid.org/0000-0002-7920-1707, Brazier, J. and Azzabi-Zouraq, I. (2017) The Use of Health State Utility Values In Decision Models. PharmacoEconomics, 35 (Suppl 1). pp. 77-88. ISSN 1170-7690

https://doi.org/10.1007/s40273-017-0550-0

THE USE OF HEALTH STATE UTILITY VALUES IN DECISION MODELS

Running header: Using HSUVs in economic models

Authors Roberta Ara[1], MSc; John Brazier[1], PhD; Ismail Azzabi Zouraq[2], MSc

[1]University of Sheffield

School of Health and Related Research (ScHARR)

Regent Court

Regent Street

Sheffield

UK

[2]Takeda Pharmaceuticals International AG

Corresponding author: Roberta Ara (r.m.ara@sheffield.ac.uk)

ABSTRACT

Methodological issues of how to use health state utility values (HSUV) in decision models arise frequently including the most appropriate evidence to use as the baseline (e.g. the baseline health state utility values associated with avoiding a particular health condition or event), how to capture changes due to adverse events and how to appropriately capture uncertainty in progressive conditions where the expected change in quality of life is likely to be monotonically decreasing over time.

As preference-based measures provide different values when collected from the same patient it is important to ensure that all HSUVs used within a single model is obtained from the same instrument where ever possible. When people enter the model without the condition of interest (e.g. primary prevention of cardiovascular disease, screening or vaccination programmes), appropriate age and gender adjusted health state utility values from people without the particular condition should be used as the baseline. General population norms may be used as a proxy if the exact condition-specific evidence is not available. Individual discrete health states should be used for serious adverse reactions to treatment and the corresponding HSUVs sourced as normal. Care should be taken to avoid double counting when capturing the effects for both less severe adverse reactions (e.g. itchy skin rash or dry cough), or more severe adverse events (e.g. fatigue in oncology) .

Transparency in reporting standards for both the justification of the evidence used and any 'adjustments' is important to increase readers' confidence that the evidence used is the most appropriate available.

Key points for decision makers

- Health state utility values (HSUVs) used in decision models should be obtained from the same data source and collected or predicted from the same preference-based measure where ever possible.
- Serious adverse reactions to treatment should be represented by individual discrete health states and care should be taken to avoid double counting for adverse reactions (e.g. itchy skin rash or dry cough)
- If no suitable evidence is identified in the literature reviews, the preferred approach would be to conduct an independent study to collect the required HSUVs. If this is not possible, evidence from people with a 'similar' condition, or a vignette study are options but both are

to be considered inferior, and a wide range of sensitivity analyses should be conducted to determine if the model results are robust to the values used.

1. Introduction

Populating decision models with health state utility values (HSUVs) is not an exact science and there are unresolved issues relating to both the evidence available (which may not always be the ideal evidence) and the practical technical problems of how to use the evidence available in decision models. These issues arise time after time and while there is some methodological research in the area, this is limited and recommendations are sparse. As a consequence, this article indicates best practice based on the limited evidence available and personal experience of the authors and is not intended to be prescriptive.

It is important to reiterate that irrespective of the specific requirements of policy makers and reimbursement agencies, it is extremely important that HSUVs within a decision model are obtained using the same preference based measure (PBM) (or are all estimated from the same valid mapping function [1]) as HSUVs obtained from different PBMs are not directly comparable due to the differences in HSUVs, even when completed by the same person [2,3,4,5,6,7]. When multiple sources of high quality relevant evidence are available, ideally a full meta-analysis would be conducted, although for many models (e.g. osteoporosis) this is not be possible due to heterogeneity in the studies [8,9]. When multiple HSUVs are available, the final choice should be clearly justified to avoid selection bias.

This article examines the use of HSUV evidence collected in clinical trials used in decision analytic models (DAM) to estimate the effectiveness of the intervention; the evidence used to represent the baseline / counterfactual health states; adjustments to available evidence to account for age, gender, adverse effects on health related quality of life (HRQoL) associated with interventions; characterising uncertainty in HSUVs; and options available when the required HSUVs are not available from a measure of HRQoL. Estimating HSUVs for comorbid conditions are discussed separately [10].

2. HSUVs collected in RCT used to estimate effectiveness

The advantages and disadvantages of efficacy randomised control trials (RCTs) as sources of data for estimating HSUVs are discussed elsewhere [11,12,13]. The main advantage stems from the internal validity arising from a RCT since the patients differ only in the treatment they receive. However, there are numerous reasons why RCTs may not provide the most appropriate evidence for estimating HSUVs for a decision model. HRQoL evidence may not be collected; the preferred preference-based measure may not be used; the sample size or numbers of events observed may be

too small to subgroup for all the individual health states within the decision model; the effectiveness evidence may be from a meta-analysis of several studies and it may be impossible to pool HRQoL evidence [14]; or the collection times of the evidence may not capture the different clinical events (e.g. it can be difficult to capture flares in inflammatory conditions). Furthermore, the patient sample and/or treatment protocols (for the new and control regimes) may not reflect those in the decision model representing the consequences for patients in the real world. The role of HSUVs from RCTs will be specific to the context.

However, irrespective of whether it is used in the decision model, if HSUVs are available from RCTs used to estimate effectiveness, then this evidence should be reported in some form to illustrate that the intervention under evaluation (or the comparator) does not have an independent detrimental effect on HRQoL over and above the effect associated with discrete clinical events. For example, pharmacological interventions prescribed to treat prostatic hyperplasia can cause sexual dysfunction [15].

3. Baseline (or counterfactual) evidence

Decision models used to assess the cost-effectiveness of interventions in health care typically assess benefits in terms of the incremental quality adjusted life-years (QALYs) accrued with avoiding a clinical event (e.g. a heart attack in people with a history of cardiovascular disease, or a hip replacement in people with osteoarthritis), a reduction in progression of a chronic condition (e.g. a reduction of pain and increase in function in people with ankylosing spondylitis), an increase in life expectancy (e.g. in people with cancer), or the alleviation or reduction in risk of contracting a particular condition (e.g. vaccinations for seasonal influenza or screening programs for cancer). Consequently, in addition to mean HSUVs obtained from people experiencing particular clinical events, procedures or conditions, DAMs also require HSUVs from people who do not experience the event, procedure or condition (see **Box 1** for examples) to determine the potential incremental gain in avoiding an event or alleviating a condition.

Box 1 Examples of different baseline HSUVs

Example 1

In a primary prevention, screening or vaccination model, most people will enter the model in a 'disease-free' health state. For example, in its simplest form, a decision model evaluating the benefits of a vaccination programme against the influenza virus would involve three health states: no influenza, influenza and death. In this instance analysts will require HSUVs for people experiencing influenza and HSUVs from the general population not currently experiencing influenza. Due to the nature of the condition, the effect on HRQoL is likely to be short-term in the vast majority of cases, with no residual long-term effects. Typically, evidence from people who do not currently have influenza will not be available, and in this instance it would be appropriate to use age and gender stratified HSUVs from the general population as evidence for the 'disease free' health state.

Example 2

Lipid lowering interventions reduce the risk of cardiovascular events and can be prescribed for primary prevention (i.e. in patients with no history of cardiovascular disease) or for secondary prevention (i.e. in patients with a history of cardiovascular disease). Consequently, in addition to the acute and long-term HSUVs from people who have experienced a particular cardiovascular event (e.g. heart attack or stroke), analysts modelling the quality adjusted life years (QALY) benefits associated with these interventions will also require HSUVs from people who have not experienced the particular event. In a primary prevention analysis (example 2a) this would be evidence from people with no history of cardiovascular disease (stratified by age and gender) while in a secondary prevention analysis (example 2b) this would be evidence from people who have a history of cardiovascular disease but no recent cardiovascular event (stratified by age and gender).

3.1 Alternative forms of evidence that can be used to depict the baseline /counterfactual

There are three forms of evidence that have been used as the baseline: a constant value of full health (i.e. EQ-5D = 1 irrespective of age or gender), age and gender stratified evidence from the general population (includes all evidence irrespective of health status), and age and gender stratified evidence from individuals who do not have the condition (or event) of interest. Assuming a baseline of full health is inappropriate as, on average, the mean HSUV is never equal to one irrespective of age or gender (**Table 1**) [16]. Using the previous examples (**Box 1**), if a heart attack is avoided in patients with existing cardiovascular disease there may still be a detrimental effect on HRQoL associated with the underlying condition (e.g. the residual long-term effects of a previous stroke or heart attack). Similarly, if influenza is avoided, given the target population (e.g. elderly and people with diabetes etc.), the recipients of the vaccinations are likely to have at least one other prevalent chronic health condition. Consequently using a baseline of full health (i.e. HSUV = 1) will overestimate the benefits of treatments (See **Box 2** for an example).

Table 1 UK EQ-5D-3L age-adjusted population norms [17,18]

Age	Male	Female
20	0.9536	0.9324
25	0.9449	0.9236
30	0.9344	0.9132
35	0.9223	0.9011
40	0.9086	0.8874
45	0.8932	0.8720
50	0.8761	0.8549
55	0.8574	0.8362
60	0.8370	0.8158
65	0.8150	0.7938
70	0.7913	0.7701
75	0.7659	0.7447
80	0.7389	0.7177

Estimated using:

$$EQ\text{-}5D = 0.9508566 + 0.021212126 \times male - 0.0002587 \times age - 0.0000332 \times age^2 \quad \text{Eqn1}$$

While evidence from people without the specific condition or event is the preferred evidence, this form of evidence is limited and can be difficult to source. It has been suggested that general population norms could be used as the baseline if evidence is not available from people without a particular health condition [17]. This is plausible if the prevalence of the condition is relatively small as revised mean HSUVs obtained when excluding the relatively small subgroup will probably not

differ substantially from the mean HSUVs from the full sample of the general population. As these values are likely to be very slightly lower than those obtained when excluding a particular population, the benefits of interventions will be underestimated (see **Box 2** for an example).

Box 2 Total and incremental QALYs associated with avoiding a heart attack

Estimating the QALYs gained when avoiding a heart attack using alternative evidence for the baseline (assuming the event occurs at the age of 50 years, and using a 50 year horizon)

Baseline trajectory	**EQ-5D**
Full health	$= 1$
Population with no history of CVD	$= 0.9454933 + 0.0256466 \times \text{male} - 0.0002213 \times \text{age} - 0.0000294 \times \text{age}^2$ [18] Eqn1
General population	$= 0.9508566 + 0.0212126 \times \text{male} - 0.0002587 \times \text{age} - 0.0000332 \times \text{age}^2$
Mean EQ-5D for heart attack	$= 0.7390$
Mean age for heart attack	$= 66.6$ years
Mean EQ-5D at age 66.6 years for pop^n with no history of CVD	$= 0.9454933 + 0.0256466 \times \text{male} - 0.0002213 \times 66.6 - 0.000294 \times 66.6^2 = 0.8260$
Multiplier	$= 0.7390 \div 0.8260 = 0.8947$ (See Appendix)
Mean EQ-5D at age 66.6 years for gen pop^n	$= 0.9508566 + 0.0212126 \times \text{male} - 0.0002587 \times 66.6 - 0.0000332 \times 66.6^2 = 0.8076$
Multiplier	$= 0.7390 \div 0.8076 = 0.9151$

Total and incremental QALYs accrued over 50 years using the alternative baseline evidence

	Baseline of full health		Baseline from people with no history of CVD		Baseline from general population	
	No heart attack	Heart attack	No heart attack	Heart attack	No heart attack	Heart attack
Total QALYs	50	36.95	39.27	35.13	38.08	34.85
Incremental QALY gain		13.05		4.14		3.23

The total QALYs accrued over a 50 year horizon are calculated for a male who does not experience a heart attack (Person A), and for a male who experiences a heart attack (Person B), using three alternative profiles for the baseline. For example, using a baseline of full health Person A would accrue 50 QALYs (50 x 1), while Person B would accrue 36.95 QALYs (50 x 0.7390), giving an incremental QALY of 13.05 (50 – 36.95).Using evidence from patients who have no history of CVD as the baseline, Patient A would accrue 39.27 QALYs ($\sum_{50}^{1} 0.9454933 + 0.0256466 \times \text{male} - 0.0002213 \times \text{age} - 0.0000294 \times \text{age}^2$) while Patient B would accrue 35.13 QALYs ($\sum_{50}^{1} 0.9454933 + 0.0256466 \times \text{male} - 0.0002213 \times \text{age} - 0.0000294 \times \text{age}^2 \times 0.8947$) giving

an incremental QALY of 4.14 (39.27 – 35.13). The QALY gains using evidence from the general population are calculated using the same methodology: Patient A accrues 38.08 QALYs ($\sum_{50}^{1} 0.9508566 + 0.0212126 \times \text{male} - 0.0002587 \times \text{age} - 0.0000332 \times \text{age}^2$), Patient B accrues 34.85 ($\sum_{50}^{1} 0.9508566 + 0.0212126 \times \text{male} - 0.0002587 \times \text{age} - 0.0000332 \times \text{age}^2 \times 0.9151$), giving an incremental QALY gain of 3.23 (38.08 – 34.85).

As can be seen, the incremental QALYs associated with avoiding a heart attack differ substantially. Using the most appropriate evidence, obtained from people with no history of CVD, the incremental QALY gain associated with avoiding a single heart attack is 4.14, compared to 13.05 when using a baseline of full health and 3.23 when using evidence from the general population.

NB: The evidence used to derive the HSUV (0.7390) was obtained from a sample who indicated they had a history of a heart attack. The sample were not sub-grouped by time since event, and in an actual cost-effectiveness model it would be important to consider this as quality of life may change over time [18,19].

Key: CVD – cardiovascular disease; QALY – quality adjusted life year

Using a case-study in primary prevention of cardiovascular disease, researchers have shown that the baseline evidence can affect the results to such an extent that these could influence a policy decision based on a given cost per QALY threshold [18]. Using a baseline of full health resulted in a higher QALY gain and thus lower cost per QALY compared to using age-adjusted profiles. This was particularly true of older aged cohorts.

3.2 Appropriateness of general population norms

One study reported that general population norms are appropriate to use as the baseline for many prevalent chronic health conditions such as cardiovascular disease, diabetes and arthritic conditions [17]. Using broadly defined self-reported health conditions and EQ-5D-3L data collected in the Health Survey for England, and comparing age-stratified scores (people without a particular condition compared to the general population), the authors reported that general population age-adjusted norms could be used if condition specific evidence was not available. Estimates of age-gender norms obtained from the Health Survey for England (see **Table 1**) are only relevant for decision models using EQ-5D-3L HSUVs obtained using the UK tariff weights (NB: As the conditions in this survey are self-reported there are arguments against the robustness of this evidence when using condition specific evidence) [20]. While there are equivalent published general population norms suitable for other settings [21], to our knowledge, evidence from individuals who do not have

particular conditions are not currently available for other preference-based measures or settings. However, many countries now conduct large national surveys which include data on HRQoL and condition at the individual level [21], and equivalent condition specific datasets could be estimated from these as and when required. If the required condition specific evidence is not available, then a range of sensitivity analyses should be performed using evidence from the general population as one end of a range of plausible values.

In summary, evidence obtained from cohorts without the health condition (or event) of interest is the ideal evidence to model the trajectory of the baseline over time. However, age and gender stratified evidence from the general population may be a good proxy (and are preferred over using a constant baseline of full health) when these data are not available.

4. Adjustments to HSUVs

Despite the substantial volume of published HSUVs, analysts are frequently faced with the dilemma of having less than ideal evidence. For example, a cross-sectional review of HRQoL evidence used in the National Institute for Health and Care Excellence (NICE) submissions (n = 46 Health Technology Assessments (HTAs) during 2004-2008) reported that over one-third (36%) of the HSUVs (n = 284) used in the associated decision models (n = 71) had been adjusted in some way [22]. The main reasons for adjustments were reported to be age, adverse events and disease progression. The authors reported a lack of clarity in reporting and it was suggested that the number was likely to be higher if actual adjustments in the decision analytic models were clearly reported in the HTAs or the associated journal articles. They also reported a wide range in methodological variation in the methods used to adjust values. Seventy-two percent of adjustments were made by either adding or subtracting a value from the original HSUV (some values appeared to be arbitrarily chosen); 18% multiplied the original HSUV by another utility value; and 10% incorporated a multivariate analysis. While intuitively, multivariate might appear to be the preferred option, there was no indication of which method would produce the most accurate results and the authors concluded that guidance was required to clarify the appropriateness of adjustments and the preferred methods for undertaking these. However, to our knowledge such guidelines do not exist as yet.

4.1 Adjustments to account for differences in age and gender

In order to capture the full benefits of treatments, many decision models assess the QALYs accrued over a lifetime-horizon. As mentioned previously, ignoring the reduction in HRQoL due to age will over-estimate the benefits of interventions. If age and gender stratified HSUVs are used as the

baseline (as in example 1 above, see **table 1**), then there is generally no reason to make any additional adjustments for these. It should be noted that the literature shows HSUVs for females are almost always lower than for males irrespective of the measure used [21]. Consequently, when modelling an entirely female population (e.g. hormone replacement therapy for menopausal women), this should be accounted for by using evidence from female cohorts. The potential equity implications of women having lower HSUVs is an issue that should be handled externally by policy decision makers.

4.2 Detrimental effects on HRQoL associated with adverse reactions to treatments

Many interventions cause adverse reactions and evidence on possible side-effects is collected routinely in clinical studies to support safety requirements for regulatory authorities such as the FDA. Some adverse reactions are acute clinical events such as the gastro-intestinal bleeds associated with long-term steroid use; others are more chronic in nature such as the persistent dry cough associated with ACE-inhibitors, or the itchy skin rash associated with barbiturates and penicillin. While acute events will generally have an obvious detrimental effect on HRQoL, less serious side-effects may be more difficult to quantify, particularly when using generic measures as they may be insensitive to the small changes in quality of life associated with these side effects.

There are often no universally accepted standard definitions that can be used to classify adverse reactions to medical interventions, and severity definitions tend to be specific to the clinical study. **Table 2** provides exemplars of two sets of definitions used [23,24]. The non-standardisation can cause uncertainty as to which adverse events should be captured in decision models. It has been suggested that HSUVs should include decrements associated with grade 3-4 adverse events [25]. Given the definitions below (grade 3: severe, grade 4: life threatening or disabling), it is likely that these would be modelled as discrete health states within the decision model. Consequently the effect on HRQoL is likely to be significant and the required HSUV would need to be collected from people experiencing the event rather than making an adjustment to a HSUV.

However, there are also occasions (such as the itchy skin rash) where less severe adverse reactions may have an independent effect on HRQoL that should be captured in the decision model. In these instances, it is important that the effect on HRQoL is not double counted. For example, if all the HSUVs in the decision model are obtained from recipients of the treatment inducing the adverse reaction, then it is inappropriate to include an additional decrement on top. Conversely, assuming that grade 3 and 4 adverse events are described by discrete health states, and the HSUVs for all

health states are obtained from the literature (i.e. from people not receiving the intervention), it may be appropriate to apply an additional decrement to these to account for any detrimental treatment effect associated with less severe adverse reactions. Both the inclusion or exclusion of less serious side effects should be justified, and sensitivity analyses performed to determine if the model results are robust to these.

One way of exploring the magnitude of effect of adverse events would be to present subgroup analysis of HSUVs from an associated clinical trial to demonstrate any potential significant differences. This could be achieved by comparing across treatment arms, or by comparing subjects who experience the adverse reaction with those who do not. Subject to sample sizes, the decision to include AEs within the economic model could be informed by analysing trial data separately for patients with and without the grade 3 or 4 AEs. The estimates of HSUVs used for this form of analysis do not necessarily need to come from the same measure as used for the HSUVs in the decision model. If there is no significant difference, then there is no reason to adjust the HSUVs in the model. If a negative independent treatment effect is observed, a threshold analysis could be performed to determine the magnitude of effect required for the incremental cost-effectiveness ratio (ICER) to go above the cost per QALY threshold.

Table 2 Exemplar definitions of adverse event severity definitions

Severity	Detailed definition
Source: Deng, 2011 [23]	
Mild	Awareness of signs and symptoms, but easily tolerated and are of minor irritant causing no loss of time from normal activities. Symptoms do not required therapy or a medical intervention; signs and symptoms are transient
Moderate	Events introduce a low level of inconvenience or concern to the participant and may interfere with daily activities, but are usually improved by simple therapeutic measures; moderate experiences may cause some interference with functioning
Severe	Events interrupt the participant's normal daily activities and generally require systemic drug therapy or other treatment; they are usually incapacitating
Source: Sibille et al, 2010 [24]	
Grade 1	Mild AE
Grade 2	Moderate AE
Grade 3	Severe AE
Grade 4	Life-threatening or disabling AE
Grade 5	Death related to AE

Key: AE – adverse events

Additional research is required to inform guidelines in this area. Table 3 provides an overview of several areas where adjustments are frequently made to HSUVs and may be used as a checklist of things to consider prior to making adjustments.

Table 3 Checklist of considerations prior to adjusting HSUVs

Adjustment	Summary of things to consider
Gender	In general, HSUVs for females are lower than for age matched males. Does the evidence sample match the gender distribution used in the model.
Age	There is a natural decline in HSUVs by age. When using a lifetime horizon, consider if you need to adjust for age. HSUVs may decline over time in chronic progressive conditions. Adjusting for age on top

	of an observed decline in HSUVs may double count the effects of age as older patients are generally more likely to be in severe health states than younger patients in progressive conditions.
Adverse events	Some adverse events may not have an independent effect on HRQoL and may be safely ignored. Some adverse events associated with interventions may have an independent effect on HRQoL. Some generic measures may be insensitive small changes in HSUVs associated with specific adverse event. Is there evidence from an alternative measure that the adverse event in question has an independent effect on HRQoL. If the HSUVs are collected from a sample receiving the intervention of interest, applying an additional detriment for the adverse event may double count the effects of the adverse event.
Time since event	The time since the event (e.g. surgical intervention, discrete health condition) may have an independent effect on HRQoL. Check the time since event in the sample with that modelled.

5. Alternative options if the required HSUVs are not available

When cost-utility analyses first started to inform policy (1990s-early 2000s), HSUVs were rarely collected in clinical studies. At that time it could be extremely difficult to identify any evidence on HRQoL for a specific condition, and it was particularly difficult to source any preference-based HSUVs from the literature. Historically, when suitable evidence was not available, analysts would

have either presented results in terms of the cost per life year (i.e. they would have ignored the effects on HRQoL), or they would have used evidence from vignettes to populate decision models.

Fortunately, the increase in the number of policy decision agencies requiring outputs in terms of the cost per QALY has had the consequence of increasing the evidence base reporting HSUVs, particularly for more prevalent health conditions. The numbers of condition specific preference-based measures (CSPBMs) and mapping studies between measures have also increased substantially in recent years (the latter predominantly after the NICE 2008 guide stated a preference for EQ-5D evidence [26]). Consequently, compared to fifteen or twenty years ago, it is now relatively rare that there is no evidence at all on HRQoL for a particular condition. However, this situation does still occur and primarily happens for less prevalent health conditions and particularly for rare orphan conditions (e.g. lysosomal acid lipase deficiency where HSUVs collected from patients with non-alcoholic steatohepatitis was used in the absence of HSUVs from the former condition) (NICE STA ongoing at time of publication [27]).

When all options have been exhausted in terms of literature reviews and mapping possibilities, the preferred approach would always be to conduct an independent study to collect the required HSUVs in the patient population of interest. However, this form of primary research cannot always be undertaken due to either fiscal or time constraints. The options in these situations are either to use HSUVs obtained from people with a 'similar' condition or to conduct a vignette study. These latter forms of evidence, and in particular evidence obtained from vignettes, should be considered as a last resort when all other possible sources have been exhausted (see Rowen et al for further information on vignettes [28]). In both cases, to avoid criticism, it is paramount to justify the approach by thoroughly documenting and reporting the literature reviews conducted. It is equally important to perform an exhaustive series of univariate sensitivity analyses together with a threshold analysis to demonstrate both the potential effect on the ICER and to identify the HSUVs required to achieve relevant cost per QALY thresholds respectively. When using evidence obtained from vignettes, any evidence from 'similar' conditions should be utilised in univariate sensitivity analysis.

6. Characterising uncertainty in HSUVs

It is now standard practice to explore the uncertainty surrounding parameters used in decision models through univariate sensitivity, threshold analyses, and probabilistic sensitivity analyses using Monte-Carlo simulations. The results, such as cost-effectiveness acceptability curves, are used to inform the decision uncertainty in the central estimate. As the ICER is a ratio, and the denominator

is the incremental QALY gain, results can sometimes be extremely sensitive to changes in HSUVs, particularly when the central estimate for the incremental QALY gain is small. Conversely, in decision models where the majority of the QALY gain is accrued from treatment induced reductions in mortality, the results from decision models can be robust to small adjustments in HSUVs.

6.1 Forms of uncertainty

There are several sources of uncertainty in HSUVs including: the preference-weights used in the tariff, the beta coefficients in mapping functions, the variance around independent mean HSUVs, the variance around correlated HSUVs, and the uncertainty around results from meta-analyses [9].

Uncertainty in preference weights: Capturing the uncertainty in preference weights requires access to individual level data and the variance covariance matrix for the algorithm used to estimate the HSUVs. This uncertainty is more likely to be characterised when conducting an economic evaluation alongside a clinical study and to our knowledge the uncertainty in the preference-weights is rarely or ever included in DAMs. One study assessed the effect of including this parameter uncertainty in the utilities associated with the treatment of ruptured aneurysms [29]. The results were compared with those obtained generated using standard bootstrapped data from a clinical trial sample (n = 1633). The authors reported little difference in the estimated uncertainty around the utilities and concluded the parameter uncertainty may be more influential on smaller valuation sets or when using value sets estimated using different methods. Research is required on additional data to determine the potential effect ignoring this uncertainty may have on resulting ICERs.

Uncertainty in mapping functions: The uncertainty in the beta coefficients from mapping functions are captured using the associated variance covariance matrix. This matrix takes into account both the uncertainty in the point estimates and the correlation between these. When using published evidence, if the required matrix is not reported (and is not available from the author), the uncertainty in the individual coefficients (but not the correlation between the sampled values) may be explored using the reported mean and standard error.

Uncertainty in mean HSUVs: Uncertainty in independent HSUVs may be characterised independently using the distributions discussed below. However, it is not appropriate to sample independently when generating values for HSUVs known to have a monotonic relationship. Examples might be where health states are defined by disease severity, whereby the HSUV for the

most severe health state would be expected to be lower than the HSUV for the least severe health state.

Uncertainty in correlated HSUVs

Where the relationship between HSUVs is known to be monotonic (e.g. when health states are defined by severity), analysts have been known to use the same random number for each HSUV in an attempt to account for possible correlation. However, this underestimates the magnitude of uncertainty and is inappropriate [30]. One appropriate method would be to use the 'Difference method' as summarised in **Box 3** (interested readers can see Stevenson et al, 2017 for more details and a worked example) [30]. This is an area where additional research looking at the possible relationship between the decrement and the observed value for the health states is required.

Box 3 Example using difference method to estimate HSUVs for correlated health states [30]

Given two health states known to have a monotonic relationship for HSUVs:

- health state A (the least severe health state),
- health state B (the most severe health state),

then

- sample HSUV for health state A as normal using a Normal distribution
- sample a decrement (expected difference in HSUV for health states A and B) using a Beta distribution
- calculate HSUV for health state B using the sampled values for health state A and the decrement

6.2 Distributional forms used to characterise uncertainty in HSUVs

HSUVs are bound by the limits of the index and are typically skewed with a minority of negative values. Despite this, in the majority of cases, the uncertainty in the mean can be explored adequately by sampling from a normal distribution using the mean and standard error. However, this method may produce implausible values outside the limits of the index when sampling for a relatively high or low HSUV in a patient level simulation. In these cases it may be better to estimate a decrement to deduct from full health. If the mean value is near the upper limit of the index, sampling from a lognormal or gamma distribution, would produce decrements on the interval (0, positive infinity). Conversely, if the mean value is near the lower limit of the index, then scaling the standard beta distribution upwards using a height parameter (lambda) will provide sample values on

the interval (0, lambda). Sampling from the proposed distribution prior to use in the decision model is good practice and histograms of sampled values should identify any anomalies in the sampled values.

One occasion where sampled values outside the limits of the index tend to be generated is when using the results of poorly executed regression analyses. Although simple linear functions obtained using ordinary least square regressions predict relatively accurate mean estimates, they can predict values outside the range of the index when covariates are involved (e.g. if the constant is greater than one and the age coefficient is negative and small, an expected value for a younger aged person could be greater than one) [1]. Any anomaly such as this will be compounded when sampling and again histograms of sampled values across all subgroups in the decision model should be examined. Similarly when using a statistical function to predict HSUVs based on a clinical measure of progression (e.g. health assessment questionnaire in rheumatoid arthritis), the sampled values at the extremes of the index / disease severity should be examined carefully.

6.3 Univariate sensitivity analyses

Univariate sensitivity analyses involving HSUVs tend to use values selected at the analysts' discretion. These could be the mean HSUV plus or minus the standard error of the mean or if this isn't available, an arbitrary value such as plus or minus 20% of the mean. It is equally important to test results from the decision model using any additional relevant evidence identified in the literature searches. This is particularly appropriate if there is no over-riding strong rational for the evidence selected for the base case, or when there is more than one published mapping function available in the literature.

6.4 Threshold analyses

Threshold analyses are informative to decision makers if the model results are very sensitive to the HSUVs used, or if the HSUVs are less than ideal. The objective is to determine the minimum (or maximum) HSUV required for the model results to be considered cost-effective at a pre-defined cost per QALY threshold. The threshold used will be setting specific and will typically be constrained to that used by the relevant policy decision makers. Examples where a threshold analyses would be used are when adjusting the HSUVs in any way, using HSUVs from a similar condition (in the absence of any HSUVs in the condition of interest), incorporating detrimental effects associated with low grade adverse reactions, and when using evidence obtained from vignettes [28].

7. Reporting standards for HSUVs used in decision models

One of the key issues commonly encountered when reviewing decision models is the poor reporting standards of the evidence used. Transparency and clarity are required when describing the HSUVs used and the process used to reach the final inclusion decision. This is particularly true when reporting the actual values used in the decision model and any associated adjustments made to these.

In addition to the variables extracted in the systematic review or mapping study, the report describing the modelling methodology should include the mean (and uncertainty) HSUVs and any adjustments made to these, any values tested in sensitivity analyses and their rational, and the distributions used in probabilistic sensitivity analyses for each health state within the decision model. Worked examples of predicted HSUVs should be provided when using the results of regression models to demonstrate both the use of the statistical model and the possible range of predicted values. The methodology used to adjust HSUVs should be clearly described and worked examples should be provided. Variances and covariance matrices used in the probabilistic sensitivity analysis should be duly reported.

8. Summary

This article has discussed many of the issues encountered when using HSUVs in economic models together with suggested methodological approaches. The most important detail is transparency in reporting standards for both the evidence used (together with justification), and any required 'adjustments' to the HSUVs used within the model. If reporting standards are high, and a full range of sensitivity analyses are conducted, this will increase reviewers' confidence that the evidence used is the most appropriate available, and policy decision makers' confidence in the uncertainty around the resulting ICER.

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, wrote the first and subsequent drafts and edited the final draft of the manuscript. IAZ made significant edits to interim and final drafts of the manuscript. JEB made substantial contributions to the initial draft and edits to the final draft of the manuscript.

Compliance with Ethical Standards

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

Conflict of interest: Ismail is employed by Takeda**.** Roberta Ara has no conflicts of interest. John Brazier has no conflicts of interest.

APPENDIX

Box A.1 Worked example illustrating the suggested methodology for Example 2 in Box 2

We need age/gender adjusted HSUVs for the health state 'no history of cardiovascular disease' but this evidence is not available in the public domain. We know that the mean UK EQ-5D-3L HSUV for a 63 year old male with a history of cardiovascular disease (but no recent cardiovascular event) is 0.78. We have a function that can be used to generate age/gender stratified norms for the UK EQ-5D-3L (Ara & Brazier, 2010):

$$EQ\text{-}5D = 0.9508566 + 0.021212126 \times male - 0.0002587 \times age - 0.0000332 \times age^2$$

The general population norms may be adjusted to account for the history of cardiovascular disease as follows:

The expected general population norm for male aged 63 is:

$$\text{HSUV} = 0.9508566 + 0.021212126 \times (1) - 0.0002587 \times 63 - 0.0000332 \times (63 \times 63)$$

$$\text{HSUV} = 0.8240$$

The multiplier required to adjust the known HSUV for a 63 year old male with a history of cardiovascular disease (but no recent cardiovascular event) is:

$$\text{Multiplier} = \text{HSUV for a 63 year old male with a history of cardiovascular disease} \div \text{expected general population norm for males aged 63}$$

$$\text{Multiplier} = 0.780 \div 0.8240$$

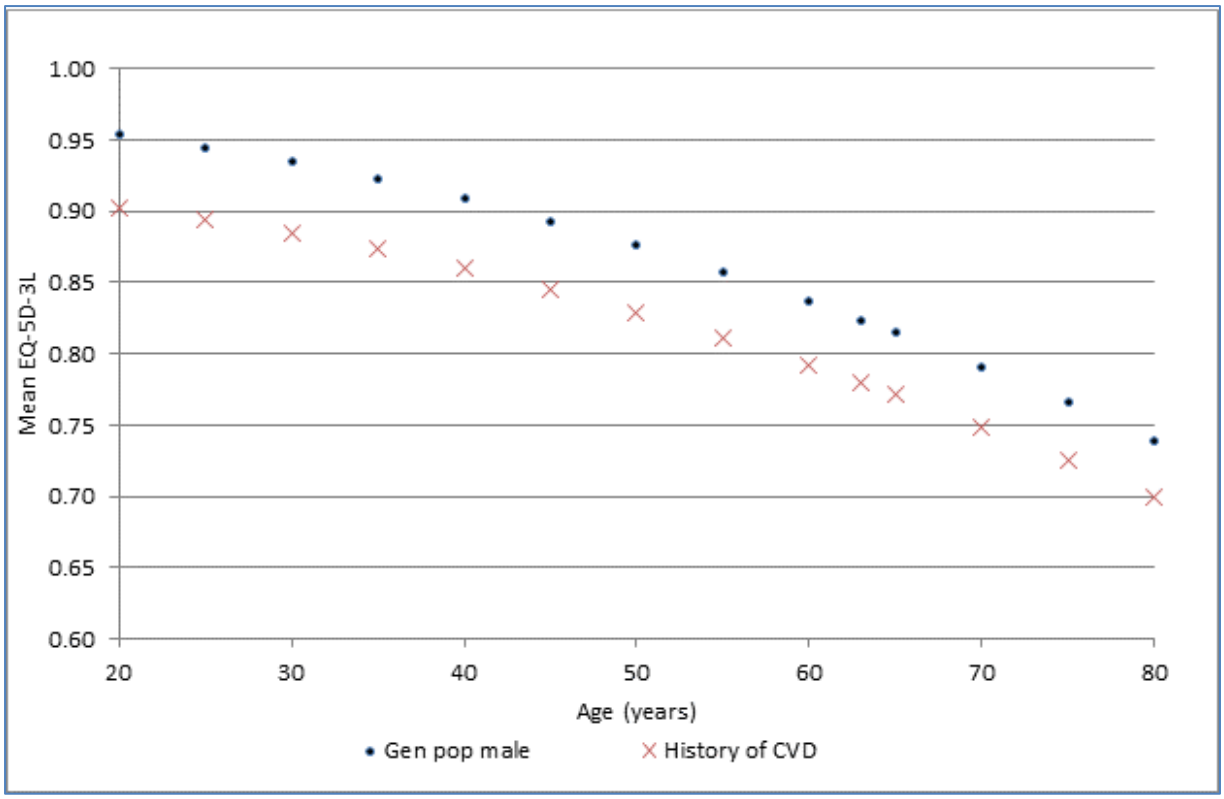
$$\text{Multiplier} = 0.9466$$

The required adjusted values for the health state 'no history of cardiovascular disease' (i.e. the baseline, are then estimated by multiplying the general population norms by the estimated multiplier. The resulting baseline HSUVs used in the model are provided in **Figure A.1**. The general population norms are plotted for illustration purposes only.

Table A.1 Age stratified health state utility values (EQ-5D)

Age	Male general population HSUVs	Age/gender HSUVs values for history of cardiovascular disease
20	0.9536	0.9027
25	0.9449	0.8944
30	0.9344	0.8845
35	0.9223	0.8731
40	0.9086	0.8601
45	0.8932	0.8455
50	0.8761	0.8294
55	0.8574	0.8116
60	0.8370	0.7923
63	0.8240	0.7800
65	0.8150	0.7715
70	0.7913	0.7490
75	0.7659	0.7250
80	0.7389	0.6994

Figure A.1 Age and gender adjusted baseline HSUVs for the health state history of cardiovascular disease



REFERENCES

[1] 04 Ara R, Rowen D, Mukuria C. The use of mapping to estimate health state utility values, Current issue Pharmacoeconomics.

[2] McLernon DJ, Dillon J. Donnan PT. Health-state utilities in liver disease: a systematic review. Med Decis Making. 2008;28(4):582–92. 4.

[3] Si L, Winzenberg TM, de Graaff B, Palmer AJ. A systematic review and meta-analysis of utility-based quality of life for osteoporosis-related conditions. Osteoporos Int. 2014;25(8):1987–97.

[4] 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.

[5] Chen J, Wong CK, McGhee SM, Pang PK, Yu WC. A comparison between the EQ-5D and the SF-6D in patients with chronic obstructive pulmonary disease (COPD). PloS one. 2014 Nov 7;9(11):e112389.

[6] Wee HL, Machin D, Loke WC, Li SC, Cheung YB, Luo N, Feeny D, Fong KY, Thumboo J. Assessing differences in utility scores: a comparison of four widely used preference-based instruments. Value in health. 2007 Jul 1;10(4):256-65.

[7] Whitehurst DG, Bryan S, Lewis M. Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. Medical Decision Making. 2011 Nov 1;31(6):E34-44.

[8] Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. Expert review of pharmacoeconomics & outcomes research. 2010 Oct 1;10(5):553-66.

[9] 04 Ara R, Brazier J, Peasgood T, Paisley S. The identification, review and synthesis of HSUVs from the literature. Current issue Pharmacoeconomics.

[10] 08 Ara R, Brazier J. Estimating health state utility values for comorbidities. Current issue Pharmacoeconomics.

[11] 06. Ara R, Brazier J. Recommended methods for the collection of HSUV evidence in clinical studies Current issue Pharmacoeconomics.

[12] 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.

[13] Brazier J, Ratcliffe J, Saloman JA, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluations. Oxford: Oxford University Press, 2017 2nd edition.

[14] 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.

[15] Stojanović NA, Ignjatović I, Đenić N, Bogdanović D. Adverse effects of pharmacological therapy of benign prostatic hyperplasia on sexual function in men. Srpski arhiv za celokupno lekarstvo. 2015;143(5-6):284-9.

[16] Fryback DG, Lawrence WG. Dollars may not buy as many QALYs as we think: A problem with defining quality of life adjustments. MDM 1997;17;276-284.

[17] 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 30;14(4):539-45.

[18] Ara R, Brazier J. Populating an economic model with health state utility values: moving towards better practice. Value in Health 2010, 13(5):509-518..

[19] Briggs AH, Bhatt DL, Scirica BM, Raz I, Johnston KM, Szabo SM, Bergenheim K, Mukherjee J, Hirshberg B, Mosenzon O. Health-related quality-of-life implications of cardiovascular events in individuals with type 2 diabetes mellitus: A subanalysis from the Saxagliptin Assessment of Vascular Outcomes Recorded in Patients with Diabetes Mellitus (SAVOR)-TIMI 53 trial. Diabetes Research and Clinical Practice. 2017 Jan 23.

[20] https://discover.ukdataservice.ac.uk/series/?sn=2000021 accessed 12 march 2017

[21] Cabases J, Janssen B, Szende A. Self-reported population health: an international perspective based on EQ-5D. Springer; 2014.

[22] Craig D, McDaid C, Fonseca T, Stock C, Woolacott N. Are adverse effects incorporated in economic models? An initial review of current practice. Health Technology Assessment. 2009; 13(62)1-181,iii.
[23] Deng. Serious Adverse Events (SAE) vs Severe Adverse Events. On Biostatistics and Clinical Trials 2011. Weblog. Available from: http://onbiostatistics.blogspot.co.uk/2011/12/serious-adverse-events-sae-vs-severe.html [Accessed 19th July 2016].
[24] Sibille M, Patat A, Caplain H, Donazzolo Y. A safety grading scale to support dose escalation and define stopping rules for healthy subject first-entry-into-man studies: some point to consider form the French Club Phase 1 working group. British Journal of Clinical Pharmacology 2010, 70(5)736-784.
[25] Ara R, Wailoo J. NICE DSU Technical Support Document 12: The use of health state utility values in decision models. London: National Health Service, 2011. Available from http://www.nicedsu.org.uk.
[26] Guide to the methods of technology appraisal, 2008. Available from www.nice.org.uk accessed 12 March 2017
[27] https://www.nice.org.uk/guidance/indevelopment/gid-lysosomalacidlipasedeficiencysebelipasealfaid737 accessed 17 March 2017 NICE (National Institute for Health and Care Excellence). Highly Specialised Technologies Evaluation: Sebelipase alfa for treating lysosomal acid lipase deficiency [ID 737]. Committee Papers. London: National Health Service, in progress (2016). Available from: https://www.nice.org.uk/guidance/GID-LYSOSOMALACIDLIPASEDEFICIENCYSEBELIPASEALFAID737/documents/committee-papers
[28] 03 Rowen D, Brazier J, Ara R, Azzabi Zouraq I, The role of condition-specific preference-based measures. Current issue Pharmacoeconomics.
[29] Gray A, Rivero-Areas O, Leal J, Dakin H, Ramos-Goni JM. How important is parameter uncertainty around the UK EQ-5D-3L value set when estimating treatment effects?. Presentation. 3rd Joint CES/HESG Meeting, Marseille, France 2012.
[30] Stevenson MD et al (personal communication: currently in progress (September 2016). Contact authors of this report or m.d.stevenson@sheffield.ac.uk directly if more information is required)