

**NICE DSU TECHNICAL SUPPORT DOCUMENT 12:
THE USE OF HEALTH STATE UTILITY VALUES
IN DECISION MODELS**

REPORT BY THE DECISION SUPPORT UNIT

July 2011

Roberta Ara, Allan Wailoo

School of Health and Related Research, University of Sheffield, UK

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent Street
Sheffield, S1 4DA

Tel (+44) (0)114 222 0734

E-mail dsuadmin@sheffield.ac.uk

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University.

The DSU is commissioned by The National Institute for Health and Clinical Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme. Please see our website for further information www.nicedsu.org.uk

ABOUT THE TECHNICAL SUPPORT DOCUMENT SERIES

The NICE Guide to the Methods of Technology Appraisalⁱ is a regularly updated document that provides an overview of the key principles and methods of health technology assessment and appraisal for use in NICE appraisals. The Methods Guide does not provide detailed advice on how to implement and apply the methods it describes. This DSU series of Technical Support Documents (TSDs) is intended to complement the Methods Guide by providing detailed information on how to implement specific methods.

The TSDs provide a review of the current state of the art in each topic area, and make clear recommendations on the implementation of methods and reporting standards where it is appropriate to do so. They aim to provide assistance to all those involved in submitting or critiquing evidence as part of NICE Technology Appraisals, whether manufacturers, assessment groups or any other stakeholder type.

We recognise that there are areas of uncertainty, controversy and rapid development. It is our intention that such areas are indicated in the TSDs. All TSDs are extensively peer reviewed prior to publication (the names of peer reviewers appear in the acknowledgements for each document). Nevertheless, the responsibility for each TSD lies with the authors and we welcome any constructive feedback on the content or suggestions for further guides.

Please be aware that whilst the DSU is funded by NICE, these documents do not constitute formal NICE guidance or policy.

Dr Allan Wailoo

Director of DSU and TSD series editor.

ⁱ National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal, 2008 (updated June 2008), London.

Acknowledgements

The DSU thanks John Brazier, Denis Fryback, Martin Hoyle, Andrew Lloyd, Paul Tappenden and the team at NICE, led by Jennifer Priaulx, for reviewing this document. The editor for the TSD series is Allan Wailoo.

The production of this document was funded by the National Institute for Health and Clinical Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

This report should be referenced as follows:

Ara, R., Wailoo, A.J. NICE DSU Technical Support Document 12: The use of health state utility values in decision models. 2011. Available from <http://www.nicedsu.org.uk>

EXECUTIVE SUMMARY

The National Institute for Health and Clinical Excellence (NICE) provides recommendations on health related quality of life (HRQoL) data used in submissions in their Guide to the Methods of Technology Appraisal. As different measures can produce different health state utility values (HSUV), the Institute state a preference for EQ-5D data to facilitate comparison across disease areas and interventions. However, inconsistencies in the methodologies used when applying HSUVs in economic models will produce discrepancies in results generated from decision analytic models, even when using the same measure, thus undermining policy decision making based on cost per quality adjusted life year (QALY) thresholds.

The objective of this Technical Support Document is to provide guidance on some of the practical issues that arise when using utility data in economic models. Specifically we look at the data used to represent the HSUVs for individuals who do not have particular health conditions (i.e. the baseline used in calculating the incremental gain), the methods used to combine HSUVs for comorbidities and the methods used to capture uncertainty in HSUVs. We describe the current evidence base in this area, provide practical advice where possible, and identify areas where current knowledge does not permit detailed guidance to be offered and additional research is warranted.

Baseline HRQoL data: When the HRQoL data in an economic model are derived from a randomised controlled trial, the HSUVs collected in the control group represent the baseline HRQoL and the HSUVs collected in the active treatment arm are used together with the baseline data to calculate the incremental QALY gain associated with treatment. Consequently the HSUVs used to inform both the baseline and the health status of the events and conditions are of equal importance. As preference-based HRQoL data are often not collected in clinical trials, these data are frequently sourced from the literature. It is inappropriate to assume the baseline is perfect health if an individual does not have a specific health condition and while ideally the baseline HSUVs would be obtained from cohorts without the health conditions or events modelled, this is not always possible. When these data are not available, using average values from the general population may be appropriate, particularly for less prevalent health conditions and those that do not have a substantial effect on HRQoL.

Adjusting/combining HSUVs: Many of the economic models submitted to the Institute use HSUVs estimated for individuals with comorbidities using data from cohorts with single specific conditions. There is currently no consensus on the most appropriate technique and the standard methods used to adjust for comorbidities (such as the multiplicative, additive or minimum methods) generate very different values. These estimated values can have substantial errors when compared to the actual values. The existing evidence base is inconclusive and additional research is required to validate emerging techniques which appear promising. In the interim period, to facilitate consistency and thus comparison of results we would recommend the multiplicative method, using adjusted baselines, is used.

Capturing uncertainty in HSUVs: Capturing uncertainty in economic model parameters is a key requirement of submissions to the Institute. Typically however, the uncertainty in the estimation of preference-based valuation weights has been ignored. In order to be able to incorporate this aspect of uncertainty, analysts require the relevant data be made available from the developers of these preference weights. We provide the required data for the Institute's preferred instrument (EQ-5D). When data are predicted using relationships between variables, the uncertainty in the coefficients should be characterised and used in the probabilistic sensitivity analyses. Capturing uncertainty in synthesised data is a developing area and until suitable techniques are identified and published, a full range of univariate sensitivity analyses should be conducted to explore the effect on results from the economic model.

CONTENTS

- 1. INTRODUCTION 8**
 - 1.1. BACKGROUND 8
 - 1.2. OBJECTIVE..... 8
- 2. BASELINE QUALITY OF LIFE..... 9**
 - 2.1. BACKGROUND 9
 - 2.2. EXISTING LITERATURE..... 10
- 3. ADJUSTING / COMBINING HEALTH STATE UTILITY VALUES.....12**
 - 3.1. BACKGROUND 12
 - 3.2. EXISTING LITERATURE..... 12
 - 3.2.1 *Combined health conditions* 12
 - 3.2.1 *Adverse events*..... 14
- 4. CAPTURING UNCERTAINTY IN HEALTH STATE UTILITY VALUES16**
 - 4.1. BACKGROUND 16
 - 4.2. EXISTING LITERATURE..... 17
 - 4.2.1. *Uncertainty in the preference-based valuation weights*..... 17
 - 4.2.2. *Uncertainty in mapping functions*..... 18
 - 4.2.3. *Synthesis of HRQoL data*..... 19
- 5. PRACTICAL ADVICE20**
 - 5.1. PRACTICAL ADVICE ON MODELLING BASELINE/COUNTERFACTUAL HSUVs 21
 - 5.1.1. *Using data from RCTs*..... 21
 - 5.1.2. *Using data from the literature* 21
 - 5.2. PRACTICAL ADVICE ON ADJUSTING OR COMBINING HSUVs 22
 - 5.2.1. *Combined health states* 22
 - 5.2.2. *Adverse events*..... 22
 - 5.3. PRACTICAL ADVICE ON CAPTURING UNCERTAINTY IN HSUVs..... 22
 - 5.3.1. *Uncertainty in the preference-based valuation weights*..... 22
 - 5.3.2. *Uncertainty in mapping functions*..... 22
 - 5.3.3. *Data synthesis* 23
 - 5.4. SUGGESTIONS FOR FUTURE RESEARCH 23
- 6. CONCLUSIONS.....23**
- 7. REFERENCES25**
- APPENDIX25**

Abbreviations and definitions

ADE	Adjusted decrement estimator
CHC	Combined health condition
CI	Confidence intervals
CVD	Cardiovascular disease
HRQoL	Health related quality of life
HSE	Health Survey for England
HSUV	Health state utility value
HTA	Health Technology Assessment
HUI3	Health Utilities Index Mark III
NICE	National Institute for Health and Clinical Excellence
OLS	Ordinary least square
QALY	Quality adjusted life year

1. INTRODUCTION

1.1. BACKGROUND

The National Institute for Health and Clinical Excellence (NICE) provide recommendations for measuring and valuing health effects in decision models submitted to the Institute through their Guide to the Methods of Technology Appraisal.¹ To facilitate comparison of results across interventions and disease areas, the Guide states a preference for EQ-5D data where possible and provides general advice on how to proceed if these data are unavailable or are considered to be inappropriate. The Methods Guide does not provide detailed instruction on how to incorporate utility values into cost-effectiveness analyses. Consistent methods to incorporate health state utility estimates into cost-effectiveness models should be applied across technology appraisals.

Although literature describing best practice in health care decision modelling is available,²⁻⁴ research specifically exploring the practicalities and issues involved when applying health-related quality of life (HRQoL) data in economic models is scarce. There are three areas where such guidance is importantly lacking: i) appropriate baseline or counterfactual data, ii) appropriate methods to combine or adjust utility measures for multiple health conditions, iii) appropriate methods for capturing the uncertainty in HSUVs used. These issues are predominantly of relevance when the analyst seeks to populate a cost-effectiveness model based on summary HSUV statistics obtained from existing literature, though some are also of relevance where the analyst has access to patient level data particularly from clinical trials.

1.2. OBJECTIVE

The objective of this Technical Support Document is to provide some insight into good practice when: modelling the HRQoL associated with not having a specific health condition (i.e. the baseline); estimating the HSUVs for comorbidities using mean values obtained from cohorts with single health conditions; and when capturing uncertainty in the HSUVs. For each of these, we describe the current evidence base, provide practical advice where possible, and identify areas where current knowledge does not permit detailed guidance to be offered. Where additional research is warranted this is highlighted.

2. BASELINE QUALITY OF LIFE

2.1. BACKGROUND

Cost-effectiveness analyses for NICE technology appraisals are often underpinned by decision analytic models. These models typically assess the benefits of health technologies in terms of the QALY gain associated with avoiding a clinical event or alleviating a particular health condition. Consequently, in addition to the HSUVs associated with the event or health condition, analysts also need the HSUVs associated with not experiencing the event or the health condition, i.e. the baseline or counterfactual values. For example, in patients with a history of cardiovascular disease (CVD), an intervention may have the potential to avoid a stroke. To model the benefits of avoiding a stroke, the analyst would ideally require a mean HSUV from a cohort of patients who have recently experienced a stroke (plus follow-up data) and a mean HSUV from a cohort who have not experienced a stroke but do have a history of CVD. Another example would be a screening programme for bowel cancer. In this instance, analysts would need the mean HSUV from a cohort with bowel cancer (plus follow-up data to capture potential changes in HRQoL over time) and a mean HSUV from a cohort who have no history of bowel cancer.

When taken as a whole, the evidence base providing condition specific HSUVs is now substantial. However, evidence about the counterfactual, namely what HSUVs patients would have experienced if they had avoided the event or condition is much more limited and it can be challenging to obtain appropriate baseline HSUVs. In the absence of these data, analysts have assumed that the alleviation of a health condition will return HRQoL to full health (i.e. EQ-5D = 1).⁵ There are several reasons why this may not be the correct approach. Using the examples in the previous paragraph, if a stroke is avoided, the patient will still have CVD which will have a detrimental effect of HRQoL; if bowel cancer is prevented, the average patient may still have at least one other prevalent health condition which will have a detrimental effect on HRQoL. Furthermore, patients who avoid an event or condition due to an intervention may be different from those who do not experience them. In addition, as lifetime horizons are frequently used to accrue the QALY gains in cost-effectiveness models, due to the increasing prevalence of comorbidities in older aged cohorts and the detrimental effect on HRQoL directly associated with age, it is reasonable to assume that the average baseline HSUVs for individuals without particular health conditions

will not be constant across the full horizon modelled. Data from the general population show that the mean HSUV for subgroups of the general population is never equal to full health irrespective of age or gender.⁶

The issue of appropriate baseline data was the subject of discussions held at a workshop used to inform the 2008 update for the NICE Methods Guide.¹ The consensus at this event was that using an arbitrary value of one is inappropriate, with the undefined caveat: “unless the nature of the disease justified its use”. Participants considered that longitudinal data from the patient group of interest would preferably be used to derive these HSUVs and it was generally agreed that adjusting for the effects of age and gender should be conducted as an absolute minimum.

2.2. EXISTING LITERATURE

It has been suggested that average HSUVs from the general population, stratified by age and gender, could be used as the baseline when condition specific data are not available.⁷ There are situations where this may be an appropriate proxy measure for patients that do not have the health condition. For example, if the specific aspects of a health condition that the technology in question seeks to alleviate or avoid are considered to be those which potentially lead to the entirety of the reduction on HRQoL associated with the general health condition then this could be an appropriate approach. Using data from the general population also makes intuitive sense as a proxy for less prevalent health conditions, or conditions that do not have a substantial effect on HRQoL as the average HSUVs derived from the general population, which consists of multiple subgroups with different conditions, would not change substantially by removing a subgroup of people who have one of these conditions. However, data from the general population may not be appropriate for more prevalent health conditions or conditions which have a substantial effect on HRQoL.

The results of a recent study suggest that while data from the general population could potentially be used as proxy baseline scores in some cases, they may not be appropriate for all.⁸ Data (n=41,174) collected in the Health Survey for England were used to compare mean EQ-5D scores for subgroups of respondents classified by self-reported health condition. Using the minimal important difference for the EQ-5D (0.074) and an assumed SD of 0.20, subgroups with greater than 64 respondents were used to detect a difference of

0.10 in mean EQ-5D scores. When comparing age-stratified scores for cohorts without a specific health condition (allowing comorbidities) with matched subgroups from the general population, the authors reported that the confidence intervals of the mean EQ-5D scores for all paired subgroups overlapped, suggesting that general population data may be appropriate when the condition specific data are not available. Conversely, when comparing mean EQ-5D scores for cohorts without a single specific health condition (but not allowing any comorbidities) with mean EQ-5D scores for similar aged respondents from the general population, the confidence intervals did not overlap ($p < 0.05$) for many of the subgroups. The authors suggested age-stratified HSUVs from respondents who had none of the prevalent health conditions could be an alternative in these instances when condition specific data are not available.

The analyses presented in this paper give EQ-5D scores for subgroups without particular health conditions. The subgroups are categorised using broad definitions which do not take into account either subtypes or severity of the condition. For example, as all individuals with a history of arthritis, rheumatism or fibrositis are included in the same category, this group will include the following cohorts: recently diagnosis of rheumatoid arthritis, severe debilitating rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis etc. In real terms, this means that although the individuals' conditions may not match the precise definition of the health condition in an economic model, as they have similar conditions, the data could potentially be more relevant for the baseline than the data from the general population.

Comparison of data across disparate sources can lead to anomalies and implausible values which may not be suitable for amalgamating within the same analysis. For example data derived from a clinical trial could indicate that a cohort with cardiovascular disease has a mean EQ-5D score of 0.75 while data from a different database could indicate that a cohort without cardiovascular disease has a mean EQ-5D score of 0.70. These problems tend to occur when combining data from different sources and can be due to: the use of different preference-based measures; differences in the definition of the condition, such as severity or disease duration; the strict inclusion/exclusion criteria used for recruitment to the studies (e.g. if individuals with comorbidities are excluded); or other factors such as differences in age or the time since a clinical event etc. As a consequence, it is extremely important that the same data source is used for all HSUVs where ever possible and the use of data from additional studies should be clearly justified. While the combination of data from different

sources can help to reduce sources of error from bias, extreme caution is required and the same degree of rigour should be applied as is required when considering clinical evidence for inclusion in meta-analyses. As a minimum, the following should be reported and compared: detailed description of the health condition including severity and duration; demographic characteristics such as age and gender; and the inclusion and exclusion criteria used in both the original studies and the review.

If condition specific data are not available, a range of sensitivity analyses should be generated with data from the general population used as one end of a range of plausible values.

3. ADJUSTING / COMBINING HEALTH STATE UTILITY VALUES

3.1. BACKGROUND

As economic models in healthcare reflect the clinical pathway followed by typical patients in general clinical practice they can become relatively complex, involving multiple health states representing the primary health condition and additional health states representing comorbidities (e.g. where an additional condition exists concurrently with the primary health condition) or adverse events associated with a treatment or intervention (e.g. nausea is a side effect of treatment given for influenza). When assessing the benefits of interventions, separate individual health states are used to define comorbidities that are relevant to the condition or treatment under evaluation.

Ideally each individual health state in the model would be populated using HSUVs derived from patients whose health condition(s) reflect the health state definitions used in the model. However, while there is a substantial evidence base providing HSUVs for individuals with single health conditions, the volume of data describing HSUVs for patients with combined health conditions (CHC) is limited.⁸

3.2. EXISTING LITERATURE

3.2.1 Combined health conditions

When HSUVs for combined health conditions are not available, these are frequently estimated using mean HSUVs obtained from cohorts with single health conditions. However, there is no consensus on the method that should be used to estimate these utility

values and different approaches can lead to vastly different estimates.⁹ Participants at the HRQoL workshop used to inform the NICE 2008 Methods Guide considered potential recommendations about approaches used to reflect the effect of comorbidities on HSUVs.¹⁰ It was agreed that these HSUVs would ideally be collected in patients with the relevant comorbidities but the participants did not agree on a preferred method for combining utility values when the relevant data were not available. It was agreed that if mean values from cohorts with the single specific conditions were used to estimate values for comorbidities, the methods and rationale should be clearly described and that this was an area for methodological research.

There are three conventional methods used to estimate the mean HSUV for a combined condition when data only exist for relevant single conditions: the additive method, which assigns a constant absolute decrement; the multiplicative method, which assigns a relative decrement, and the minimum method, which assumes no additional decrement over that observed for the condition with the lowest HSUV. Two alternative approaches recently proposed include: the adjusted decrement estimator (ADE), which is a variation of the minimum method; and a simple linear model, which incorporates terms representing the additive, multiplicative and minimum methods.^{11,12} Additional details of the alternative methods are provided in Appendix 1.

A recent review of existing empirical literature identified 11 studies exploring the accuracy of and/or comparing the performance of the alternative methods currently used to estimate HSUVs for comorbidities.⁸ Three of the studies used individual level data directly elicited from patients (n=50 to 207) using either time trade-off or standard gamble.¹²⁻¹⁴ Eight studies used preference-based HSUVs obtained from generic questionnaires (EQ-5D = 4, SF-6D = 3, HUI3 = 1) evaluating data (range 5,224 to 131,535 respondents) from large surveys. In general these data were sub-grouped into very broadly defined disease areas such as “cancer (neoplasm) including lumps, mass”, and “musculoskeletal or arthritis/rheumatism/fibrositis”. There were substantial differences in the number of HSUVs estimated for the combined conditions with estimated HSUVs ranging from just 3 to 760. Two studies evaluated just one of the methods while the others compared results generated using two, three, or more.

The authors of the review found the accuracy of the methods differed across the range of actual utilities estimated which influenced analysts conclusions. Second, the majority of studies assessed the methods by comparing the mean errors in estimated values which masked bias when the size of the errors was correlated to the values being estimated. Finally, the values assigned to normal health influenced the accuracy of the methods with the errors in predicted values being larger when the decrement used was calculated using full health (i.e. EQ-5D = 1) as the starting point. They concluded that although there is no unequivocal evidence for supporting one particular method, the combination linear model appeared to give more accurate results in the studies they reviewed. However, they cautioned these models have not been validated in external data and additional research is required before this method could be recommended. Of the other methods compared, although all estimated some values with substantial errors, the multiplicative appears to be the most accurate overall. An additional study published after this review advocated the multiplicative method when comparing the three standard methods on SF-6D utilities.¹⁵

The uncertainty in the values estimated using the different techniques has not been studied, and very little research exploring accuracy in methods when estimating HSUVs for more than two simultaneous conditions has been conducted.

3.2.1 Adverse events

When considering adverse events, it is important to distinguish between acute events and chronic sequelae and the inclusion of decrements on HRQoL associated with grade 3-4 adverse events is particularly important. Conversely, care is required to ensure the decrements associated with grade 1-2 adverse events are not double counted as the cohort used for the main HSUVs may have included a proportion of patients who had experienced these adverse events. As in the preceding section, treating the decrement associated with the adverse event as a constant value may be inappropriate.

A review commissioned by the National Institute for Health Research Health Technology Assessment (HTA) programme examined current practice when incorporating adverse events in economic models described in HTA reports published between 2004 and 2007.¹⁶ Forty-seven of the 80 studies reviewed were assessments conducted to inform NICE appraisals. The authors concluded that adverse events are often not taken into account and when they are included in the economic model, the reporting of the evidence used is poor.

As the objective of the review was not to identify the methodologies used in applying HSUVs associated with adverse events, the techniques used were not reported. Hence it is not clear if any additional adjustments were made to utilities to reflect adverse events. The authors stated there was a need for more transparency as very few reports made an explicit statement that the utility valuation captured adverse effects. This is particularly relevant when utility data are sourced from the literature and are not obtained from recipients of the intervention under evaluation. They recommended a clear justification should be provided for the non inclusion of adverse effects and advocated separate sections on adverse effects in both clinical and cost-effectiveness sections of HTA reports, including an explicit report of how adverse effects are considered in the decision model. They noted that the links between the clinical review and the modelling components were not strong for adverse effects with only 21% of modellers relying solely on the accompanying review for the required data. In addition, they reported it was apparent that non-systematically derived literature-based data were the most commonly used. They suggested more rigorous methods need to be adhered to and recommended research exploring the best approach to ensure that any adverse effects of interventions are captured within the utility was required.

Authors of a recent cross-sectional review of HRQoL data used in Health Technology Appraisals (n=46) submitted to NICE during the period 2004-2008 found almost half of HSUVs (43%) used in the 71 models in the submissions did not meet the reference case.¹⁷ Over one third (36%) of the 284 utility values used in the 71 models, were adjusted in some way. Six for patients' age, 1 for gender, 13 for adverse events, 16 to account for disease progression, 2 to include a carer's HRQoL, 3 for other reasons and 1 reason was unclear. Adjustments were made by either adding or subtracting a value (72%) from the original HSUVs; multiplying by a utility weight (18%); or incorporating a multivariate analysis (10%). It was noted that the total number of adjustments could be higher if the adjustments made in the electronic models were not reported in the accompanying reports or related publications. The authors reported a wide range of methodological variation in the use of utility values and a lack of clarity in the reporting of detailed methods used in the submissions. They concluded further guidance is required to clarify the appropriateness of adjusting values and the preferred methods for undertaking these adjustments.

4. CAPTURING UNCERTAINTY IN HEALTH STATE UTILITY VALUES

4.1. BACKGROUND

Participants at the NICE HRQoL workshop used to inform the current Guide to Technology Appraisals considered that uncertainty around HSUVs is usually under-reported in health economic submissions and that frequently only mean values are reported.¹⁰

Referring to uncertainty around mean health inputs to the model i.e. parameter precision, the NICE Methods Guide states:

“The mean value, distribution around the mean, and the source and rationale for the supporting evidence should be clearly described for each parameter included in the model. The distributions chosen for probabilistic sensitivity analysis should not be arbitrarily chosen, but chosen to represent the available evidence on the parameter of interest, and their use should be justified. Formal elicitation methods are available if there is a lack of data to inform the mean value and associated distribution of a parameter. If there are alternative plausible distributions that could be used to represent uncertainty in parameter values, this should be explored by separate probabilistic analyses of these scenarios”.¹

Referring to the characterisation of potential bias and uncertainty in the inputs to decision-analytic models, the NICE Method Guide states:

“The implications of different estimates of key parameters must be reflected in sensitivity analyses (for example, through the inclusion of alternative scenarios). Inputs must be fully justified and uncertainty explored by sensitivity analysis using alternative input values.”¹

HRQoL data, and in particular the preference-based data preferred by NICE (i.e. EQ-5D), are not normally distributed. They are typically skewed, bimodal or trimodal, are bounded by the limits of the index and often involve negative values.¹⁸ However, in the majority of cases, the uncertainty in the mean can be adequately described by sampling from a normal distribution. There will be exceptions to this, for example when sampling for a patient level simulation model using data from a cohort which has a wide variation in HRQoL scores and a relatively low or high mean score. In these instances, an alternative approach would be to

describe the utility values as decrements from full health (i.e. 1 minus the HSUV)¹⁹ and then sample from a log normal or gamma distribution giving a sampled utility decrement on the interval (0, positive infinity). If a lower constraint is required (e.g. -0.594 for the UK EQ-5D index), the standard beta distribution could be scaled upwards using a height parameter (λ) producing a distribution on a (0, λ) scale.

There are several issues in capturing uncertainty in HSUVs that are not covered in the NICE Methods Guide including: uncertainty in preference-based valuation weights; uncertainty in mapping functions, and data synthesis.

4.2. EXISTING LITERATURE

4.2.1. Uncertainty in the preference-based valuation weights

Currently, the full uncertainty associated with HSUVs is rarely captured in economic models as the uncertainty relating to the original construction of the preference-based weights is generally not characterised in the process. HRQoL instruments such as the EQ-5D, SF-6D, and Health Utilities Index 3 (HUI3) consist of a series of statements asking respondents to value their own health. For example the EQ-5D consists of five statements, each of which has three response choices, giving a total of 243 health states. As it is not feasible to value all health states generated by the instruments, a selection of health states are typically valued. In general, the states valued for these kinds of instruments will have been selected based on designs such as orthogonality, to ensure that the data is gathered efficiently given the model specification that will be used. The statistical models fitted to health states valued will consist of one or more parameter estimates (preference weights) and for each estimate there will be a measure of uncertainty, e.g. a standard error. The statistical models are subject to model parameter, model structure and model process uncertainties. When the preference weights (point estimates of the model parameters) are applied to other datasets to predict utility values this uncertainty is typically ignored. The characterisation of parameter uncertainty should be incorporated in all submissions to NICE¹ and it is possible to incorporate the uncertainty surrounding the preference-weights in probabilistic sensitivity analyses using the associated variance covariance matrices (for the EQ-5D matrix see <http://www.nicedsu.org.uk/Utilities-TSD-series%282391676%29.htm>). For completion, we recommend that the variance covariance matrices for all the preference-based indices are

made available to enable all the uncertainty surrounding health state utility values to be captured in the economic models.

Decision models generally incorporate multiple health states depicting changes in health conditions such as disease progression, adverse events or distinct events such as heart attacks and strokes. Each of these health states can have a different HSUV and correlations between these values should be characterised in the probabilistic sensitivity analyses using multivariate distributions where possible. Alternative approaches are currently being explored and the resulting recommendations will be a useful reference for analysts. In addition, a recent publication suggests the standard error of measurement for a number of leading health utilities varies depending where along the health continuum the measurement is made.²⁰ These last two are examples of the growing volume of work in this area and the literature should be continually reviewed to take account of emerging evidence.

4.2.2. Uncertainty in mapping functions

The NICE Methods Guide states that EQ-5D utility data can be estimated using mapping functions generated from HRQoL or clinical variables.^{1,17} Potentially as a result of this, there has been a growth in publications describing research in this area and there are now numerous articles providing models that can be used to estimate the required preference based data from other HRQoL data.^{17,21} However, very few authors provide the data required to incorporate uncertainty in probabilistic analyses such as the variance covariance matrix.

Over a quarter (19/71) of recent submissions to NICE used mapping functions to predict utility values and although not stated explicitly, we assume that the majority of these functions were derived using patient level data relevant to the particular submission as opposed to published algorithms.¹⁷ It was not clear how many, if any, incorporated the uncertainty associated with the point estimates from the mapping algorithm in the probabilistic sensitivity analyses. The authors of the review noted that the descriptions of the methods and any validation exercises varied considerably and recommended further methodological guidance in this area.

4.2.3. Synthesis of HRQoL data

Economic models generally involve multiple health states, for example health states can be defined to reflect improvements or deterioration in health over time. While HSUVs can be sourced from the clinical studies used to inform the effectiveness of the intervention under evaluation, decision-analytic models typically incorporate HSUVs from more than one source (see Utilities TSD 9).²² Problems can arise when literature searches produce a wide variation in values for a given health condition and synthesis of these data can be limited due to heterogeneity in the studies and data identified. The NICE Methods Guide states there should be “transparent and reproducible synthesis of all relevant evidence on health effects” and that “evidence about the extent of correlation between individual parameters should be carefully considered and reflected in the probabilistic analysis. Assumptions made about the correlations should be clearly presented.”¹

Authors of a review describing the use of evidence in all NICE technology assessment reports (n=41) published between January 2003 and July 2006 reported that the results of HRQoL systematic reviews were never used in the cost-utility analyses because preference-based HRQoL measures are rarely included in clinical trials.¹⁶ They concluded that the methods used to search for evidence were not appropriate in many cases, and the method chosen to summarise the data (i.e. median as opposed to mean) may inhibit assessments of economic benefit and that relevant data for estimating QALYs were not contained in the reviews. The authors suggested that the reluctance of systematic reviewers to pool data due to heterogeneity in clinical studies can impede the production of pooled estimates and cautioned that alternative approaches such as using a subset of studies or using expert opinion could be even more unsatisfactory.

Participants at the HRQoL workshop used to inform the updated Methods Guide considered that while it is theoretically possible to synthesise HSUVs across studies, due to heterogeneity between studies this is difficult in practice. Compliance with the reference case criteria (i.e. EQ-5D data) was generally considered to be a minimum standard for synthesis. It was noted by the participants that the meta-analysis of HUSVs is an emerging methodological area and methods are not yet fully developed (see TSD9).²²

However, meta-analysis of HSUVs is an emerging methodology and the appropriate methods for undertaking this type of synthesis are currently unclear. Conditions where

meta-analysis of HSUVs have been conducted include breast cancer,²⁴ HIV/AIDS,²⁵ liver disease,²⁶ osteoporosis,²⁷ prostate cancer,²⁸ stroke,²⁹ and renal replacement therapy.³⁰ Synthesis techniques ranged from a crude pooling of the data²⁷ to meta-regressions using pooled ordinary least squares (OLS).²⁴

- *Crude pooling*: A review of 27 studies describing utility values associated with osteoporosis conditions reported substantial variations in reported values.²⁷ For example when looking at the six studies reporting EQ-5D values (UK tariff) related to hip fractures, the average EQ-5D scores are estimated to range from 0.58 to 0.84 (pooled estimate 0.756) before the fracture, and from 0.46 to 0.73 (pooled estimate 0.580) at 12 months after the fracture (values estimated from Figure 3). The authors recommended crude pooling of the data (weighted by the inverse of the variance or sample size) as more sophisticated methods of synthesis were limited by insufficient number of studies and the wide range of valuation methods used within the studies.²²
- *Meta-regressions using OLS*: A systematic review providing a pooled estimate for identifiable health states for breast cancer included 49 relevant papers each of which contributed between 1 and 36 HSUVs covering six categories ranging from screening-related states to early or metastatic breast cancer.²⁴ In total, these gave 230 HSUVs for early breast cancer states and 117 HSUVs for metastatic breast cancer. These data were combined using ordinary least squares regressions with utilities clustered within study group and weighted by both number of respondents and inverse of the variance of each HSUV. Covariates included disease state, utility assessment method and other features of study design. Despite the substantial number of HSUVs, it was not possible to generate a definitive list of EQ-5D data for use in future economic evaluations due to the complexity of the health states involved and the variety of methods used to obtain values.

5. PRACTICAL ADVICE

It is difficult to provide detailed evidence-based practical advice due to the scarcity of robust research in this area. While the following section provides practical advice within the limitations of the existing evidence base, this advice should be updated as results of research in this area become available.

5.1. PRACTICAL ADVICE ON MODELLING BASELINE/COUNTERFACTUAL HSUVs

5.1.1. Using data from RCTs

For decision-analytic models incorporating HSUVs from randomised controlled trials, data from the control group should be used as the baseline and data from the active treatment arm would then be used to calculate the incremental QALY gain accrued through the intervention. When extrapolating beyond the duration of the clinical trial, for example when using a lifetime horizon, analysts should endeavour to supplement the HSUVs used to account for potential changes due to factors such as age and increasing numbers of comorbidities, for example by using data from the general population as the baseline. A clear description of the additional data used including the patient demographics should be reported and compared with corresponding data from the clinical study together with a full description of any assumptions used. Where possible, uncertainty in all HSUVs should be incorporated in the probabilistic analyses and it is sometimes also helpful to conduct univariate sensitivity analyses to assess the independent effect the HSUVs have on the results generated.

5.1.2. Using data from the literature

For decision-analytic models incorporating HSUVs derived from alternative sources, it is not appropriate to use a baseline of full health (i.e. EQ-5D = 1) to calculate the HRQoL benefits of treatments. Where possible, data obtained from people without the specific health condition or event under consideration should be used to inform the baseline. If these data are not available, data from the general population may be appropriate for some analyses depending on the definitions of health states within the model. For other analyses, data from people with none of the most prevalent health conditions may be used. A clear description of the data used including the patient demographics should be reported and compared with corresponding data from the clinical study together with a full description of any assumptions used. Where possible, uncertainty in all HSUVs should be incorporated in the probabilistic analyses and a full range of sensitivity analyses using upper and lower 95% confidence intervals should be conducted to explore if economic results are sensitive to changes in HSUVs.

5.2. PRACTICAL ADVICE ON ADJUSTING OR COMBINING HSUVS

5.2.1. Combined health states

When HSUVs from cohorts with combined health conditions are not available, based on the current evidence, the multiplicative method should be used to combine the data derived from subgroups with the single health conditions (Appendix). The multiplier used to combine these data should be estimated using age-adjusted data as a minimum to increase accuracy in the estimated values.⁸

5.2.2. Adverse events

The description of cost-effectiveness analyses submitted to NICE should provide clear justification for either inclusion or exclusion of disutilities for adverse events in the decision-analytic model. Where the adverse events are known to affect HRQoL they should be treated in the same way as the associated costs and included in the model. Authors should provide a detailed description of parameter values and methods used to source evidence.²² If the data require adjustments, the recommendations above for combined health states should be followed and a full range of sensitivity analyses conducted to determine the effect on results when varying the input parameters. Where possible, uncertainty in any adjustments should be incorporated into the probabilistic sensitivity analyses. The means by which uncertainty is characterised, including choices of statistical distributions and correlations, should be fully documented and justified.

5.3. PRACTICAL ADVICE ON CAPTURING UNCERTAINTY IN HSUVS

5.3.1. Uncertainty in the preference-based valuation weights

Uncertainty in the preference-based valuation weights should be incorporated in the probabilistic sensitivity analyses using the variance covariance matrix associated with the preference-based weights. The variance-covariance matrix for the EQ-5D preference-based algorithm is available (www.nicedsu.org.uk/Utilities-TSD-series) and attempts should be made to obtain similar data from the corresponding author if these are not available.

5.3.2. Uncertainty in mapping functions

Additional research is currently ongoing into the most appropriate methodologies for mapping functions using HRQoL data.¹⁷ Until results of research in this area are published, analysts developing mapping functions for use in economic models should ensure that the

outputs are reported in such a way as to facilitate uncertainty analyses in the models i.e. variance covariance matrices should be reported and these should be incorporated in probabilistic sensitivity analyses.

5.3.3. Data synthesis

As data synthesis of HSUVs is an emerging methodological area the evidence base should be reviewed periodically. At present, it is premature to recommend anything more than a weighted average or pooling in the synthesis of HSUVs used in economic models.²² A full range of sensitivity analyses should be conducted to determine the effect on results when varying the input parameters.

5.4. SUGGESTIONS FOR FUTURE RESEARCH

This is an area where a substantial amount of research is required to support the practical advice proposed in this report. Research is required in the following areas:

- Analyses of longitudinal data describing potential changes in HSUVs for subgroups of patients with specific health conditions.
- Analyses exploring baseline for the counterfactual health states in analytic models to enable more precise calculations of the incremental health benefits of treatment
- Empirical research on the most appropriate method for adjusting data to reflect comorbidities and/or adverse events.
- Primary studies collecting data for acute events are required.
- Analyses to determine the class/type and duration of adverse event that should be incorporated in economic models.
- Additional research assessing the existing methods for combining HSUVs.
- Additional research into methods used to meta-analyse HSUVs taking into account the requirements of economic models.

6. CONCLUSIONS

There is evidence that the methodologies used to apply HSUVs can make a substantial difference to the results generated from cost-effectiveness models which will undermine consistent decision making. However, this paper has identified that a substantial volume of research is required before definitive detailed evidence-based practical advice on using

HSUVs in economic models can be offered to analysts submitting economic models to NICE. One theme that was apparent in much of the evidence reviewed was a lack of clarity and transparency in current reports of methodologies used.

7. REFERENCES

1. National Institute for Health and Clinical Excellence Guide to the methods of technology appraisal (updated June 2008). 2008.
2. Briggs, A.H. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 2000; 17(5):479-500.
3. Willan, A.R., Briggs, A.H., Ebooks Corporation. Statistical analysis of cost-effectiveness data. John Wiley, 2006.
4. Gold, M.R. Cost-effectiveness in health and medicine. Oxford University Press, USA, 1996.
5. Ward, S., Lloyd Jones, M., Pandor, A., Holmes, M., Ara, R., Ryan, A. et al. A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technol Assess* 2007; 11(14):1-160.
6. Dolan, P. The time trade-off method: results from a general population study. *Health Economics* 1996; 5(2):141-154.
7. Fryback, D.G., Lawrence, W.F. Dollars may not buy as many QALYs as we think. *Medical Decision Making* 1997; 17(3):276.
8. Ara, R., Brazier, J.E. 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 2011; 14 (4): 539-545.
9. Ara, R., Wailoo, A.J. Estimating health state utility values for comorbid health conditions: a synopsis of the current evidence base. 2010; available from <http://eprints.whiterose.ac.uk/>
10. Longworth, L.J., Chandiwana, D., Kandaswamy, P., Tucker, H., Hay, N., Chong, C.A.K.Y. Report to the NICE Methods Review Working Party: Key issues arising from workshop on health-related utility measurement. 2011.
11. Hu, B., Fu, A.Z. Predicting utility for joint health states: a general framework and a new nonparametric estimator. *Medical Decision Making* 2010; 30(5):E29.
12. Basu, A., Dale, W., Elstein, A., Meltzer, D. A linear index for predicting joint health states utilities from single health states utilities. *Health Economics* 2009; 18(4):403-419.
13. Esnaolaa, N.F., Doherty, D.A., Johnson, M.L., Ellis, V.O., Miller, A.R., Janjan, N.A. et al. Comparison of additive and multiplicative utility predictions to predict the utilities of combined health states. Abstracts from the 23rd Annual Meeting of the Society of Medical Decision Making. *Medical Decision Making* 2001; 21:517.
14. Dale, W., Basu, A., Elstein, A., Meltzer, D. Predicting utility ratings for joint health states from single health states in prostate cancer: empirical testing of 3 alternative theories. *Medical Decision Making* 2008; 28(1):102.

15. Hanmer, J., Vanness, D., Gangnon, R., Palta, M., Fryback, D.G. Three methods tested to model SF-6D health utilities for health states involving comorbidity/co-occurring conditions. *Journal of Clinical Epidemiology* 2010; 63(3):331-441.
16. Drummond, M.F., Iglesias, C.P., Cooper, N.J. Systematic reviews and economic evaluations conducted for the National Institute for Health and Clinical Excellence in the United Kingdom: A game of two halves? *International Journal of Technology Assessment in Health Care* 2008; 24(02):146-150.
17. Longworth, L., Rowen, D. NICE DSU Technical Support Document 10: The use of mapping methods to estimate health state utility values. 2011; available from <http://www.nicedsu.org.uk>
18. Hernandez Alava, H., Wailoo, A.J., Ara, R. Tails from the Peak District: adjusted censored mixture models of EQ-5D health state utility values. 2010; available from Available at: <http://eprints.whiterose.ac.uk>
19. Brazier, J., Ratcliffe, J. Measuring and valuing health benefits for economic evaluation. Oxford University Press, USA, 2007.
20. Palta, M., Chen, H., Kaplan, R.M., Feeny, D., Cherepanov, D., Fryback, D.G. Standard error of measurement of 5 health utility indexes across the range of health for use in estimating reliability and responsiveness. *Medical Decision Making* 2011; 31(2):260-269.
21. Brazier, J.E., Yang, Y., Tsuchiya, A., Rowen, D.L. 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.
22. Papaioannou, D., Brazier, J.E., Paisley, S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. 2011; available from <http://www.nicedsu.org.uk>
23. Barbieri, M., Drummond, M.F., Puig Junoy, J., Casado Gomez, M.A., Blasco Segura, P.B., Poveda Andres, J.L. A critical appraisal of pharmacoeconomic studies comparing TNF alpha antagonists for the rheumatoid arthritis treatment. *Expert Rev Pharmacoecon Outcomes Res* 2007; 7:613-626.
24. Peasgood, T., Ward, S.E., Brazier, J.E. A review and meta-analysis of health state utility values in breast cancer. 2010; available from Available from <http://eprints.whiterose.ac.uk>
25. Tengs, T.O., Lin, T.H. A meta-analysis of utility estimates for HIV/AIDS. *Medical Decision Making* 2002; 22(6):475.
26. McLernon, D.J., Dillon, J., Donnan, P.T. Systematic Review: Health-State Utilities in Liver Disease: A Systematic Review. *Medical Decision Making* 2008; 28(4):582.
27. Peasgood, T., Herrmann, K., Kanis, J.A., Brazier, J.E. An updated systematic review of Health State Utility Values for osteoporosis related conditions. *Osteoporosis International* 2009; 20(6):853-868.

28. Bremner, K.E., Chong, C.A.K.Y., Tomlinson, G., Alibhai, S.M.H., Krahn, M.D. A review and meta-analysis of prostate cancer utilities. *Medical Decision Making* 2007; 27(3):288.
29. Tengs, T.O., Lin, T.H. A meta-analysis of quality-of-life estimates for stroke. *Pharmacoeconomics* 2003; 21(3):191-200.
30. Liem, Y.S., Bosch, J.L., Myriam Hunink, M.G. Preference Based Quality of Life of Patients on Renal Replacement Therapy: A Systematic Review and Meta Analysis. *Value in Health* 2008; 11(4):733-741.
31. Keeney, R.L., Raiffa, H. Decision making with multiple objectives. 1976. Wiley, New York.
32. Keeney, R.L., Raiffa, H. The multiplicative utility function. *Decision making with multiple objectives: preferences and value tradeoffs*. Cambridge, UK: Cambridge University Press; 1993.
33. Tversky, A., Kahneman, D. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 1992; 5(4):297-323.

APPENDIX

Methods Used To Estimate HSUVs For Comorbid Health Conditions

The techniques described below use mean HSUVs from cohorts with single health conditions to estimate mean HSUVs for cohorts with CHCs. There are three main methods used to estimate the utility value for a combined health state when data only exist for relevant single health states. These can be termed the “additive”, “multiplicative” and “minimum” approaches. Alternatives recently proposed include: the adjusted decrement estimator (ADE) which is a variation of the minimum method, and a simple linear model, based on multi-attribute utility theory and prospect theory, which incorporates terms representing the additive, multiplicative and minimum methods.^{11,12}

Given two health conditions, condition A and condition B, there are four possible combinations of these conditions: individuals have condition A but not condition B, individuals have condition B but not condition A, individuals have both condition A and condition B; individuals do not have either condition A or condition B. The HSUVs associated with these four alternatives are defined as: U_A , U_B , $U_{A,B}$, and $U_{nA,nB}$.

Additive method. The additive method assumes a constant absolute decrement relative to the baseline and the estimated HSUV for the additive CHC is calculated using:

$$U_{A,B}^{add} = U_{nA,nB} - ((U_{nA} - U_A) + (U_{nB} - U_B)) \quad (\text{Eqn 1})$$

where the superscript “add” denotes the additive method.

If a baseline of perfect health is used, the additive method can be calculated using:

$$U_{A,B}^{add} = U_A + U_B - 1 \quad (\text{Eqn 2})$$

Multiplicative method. The multiplicative method assumes a constant proportional decrement relative to the baseline and the estimated HSUV is calculated using:

$$U_{A,B}^{Mult} = U_{nA,nB} \cdot \left(\frac{U_A}{U_{nA}} \right) \cdot \left(\frac{U_B}{U_{nB}} \right) \quad (\text{Eqn 3})$$

where the superscript “Mult” denotes the multiplicative method.

If a baseline of perfect health is used, the multiplicative method can be calculated using:

$$U_{A,B}^{mult} = U_A \cdot U_B \quad (\text{Eqn 4})$$

Minimum method. The minimum method assumes the decrement on HRQoL associated with a comorbidity is equal to the maximum decrement attributable to the individual single health conditions, and the estimated HSUV is calculated using:

$$U_{A,B}^{\min} = \min(U_{nA,nB}, U_A, U_B) \quad (\text{Eqn 5})$$

where the superscript “min” denotes the minimum method.

If a baseline of perfect health is used, the minimum method can be calculated using:

$$U_{A,B}^{\min} = \min(U_A, U_B) \quad (\text{Eqn 6})$$

Adjusted decrement estimator. The adjusted decrement estimator (ADE) has recently been proposed as an alternative method to estimate HSUVs for CHCs.¹¹ This estimator is a variation of the minimum method and assumes the estimated HSUV for the CHC has an upper bound equal to the minimum of the HSUVs from the two single health conditions. The proposed method is described by:

$$U_{A,B}^{ADE} = \min(U_A, U_B) - \min(U_A, U_B) \cdot (1 - U_A) \cdot (1 - U_B) \quad (\text{Eqn 7})$$

where the superscript “ADE” denotes the adjusted decrement estimator.

Combination model. Basu *et al.* recently proposed a simple linear model which incorporates terms representing the additive, multiplicative and minimum methods.¹² The model is formulated from a) an adaptation of work originally presented by Keeny and Raiffa which was based on decision theory and multi-attribute utility functions,^{31,32} and b) a prospect theory that proposes the value function is convex for losses with a marginal rate of decrement in value with increasing losses, as presented by Tversky and Kahneman (1992).³³ The model is defined by:

$$U_{A,B}^{\text{comb}} = 1 - \left(\beta_0 + \beta_1 \cdot \min((1 - U_A), (1 - U_B)) + \beta_2 \cdot \max((1 - U_A), (1 - U_B)) \right) + \beta_3 \cdot (1 - U_A) \cdot (1 - U_B) + \varepsilon \quad (\text{Eqn 8})$$

where the superscript “comb” denotes the combination model, ε the residual and the beta coefficients are obtained using ordinary least square regressions. Equation 8 uses a baseline of perfect health. Using an adjusted baseline, the combination model can be defined by:

$$U_{A,B}^{\text{comb}} = \beta_0 + \beta_1 \cdot \min((U_{nA} - U_A), (U_{nb} - U_B)) + \beta_2 \cdot \max((U_{nA} - U_A), (U_{nb} - U_B)) + \beta_3 \cdot \left(U_{nA,nB} \cdot \frac{U_A}{U_{nA}} \cdot \frac{U_B}{U_{nb}} \right) + \varepsilon$$

(Eqn 9)

The combination model reduces to the three traditional methods under the following conditions:

When $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$ and $\beta_3 = 0$, then Eqn 8 collapses to Eqn 2 (additive method)

When $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$ and $\beta_3 = -1$, then Eqn 8 collapses to Eqn 4 (multiplicative method)

When $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 0$ and $\beta_3 = 0$, then Eqn 8 collapses to Eqn 6 (minimum method)