NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Report to the Methods Review Working Party

Key issues arising from workshop on measuring and valuing health effects

This report is written by members of the Institute's team of analysts. It is intended to highlight key issues arising from discussions at the workshop on structured decision making. It is not intended to provide a detailed account of all comments expressed at the workshop. The report has been written independently of the people who attended the workshop.

The report is circulated to the members of the Method's Review Working Party, the group responsible for updating the guide. For further details regarding the update of the Guide to the Methods of Technology Appraisal please visit the NICE website at http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/GuideToMethodsTA201112.jsp.

1 Summary

There was no indication during the workshop discussions that it is necessary to deviate from the EQ-5D as the preferred utility measure in the reference case. A few participants thought that there should be no exceptions to the use of the EQ-5D, but the majority agreed that there are circumstances where other generic measures or mapping of condition-specific measures could be used. The following topics were addressed:

- **When is EQ-5D not appropriate?** Participants felt that the current Methods Guide does not adequately describe when the EQ-5D is not appropriate, and what criteria determine which preferred measure should

be presented instead. Where alternative measures are presented, there was a preference for these to be presented alongside the EQ-5D wherever possible, and also for any alternative measure to use preference-based valuation methods, ideally the time-trade-off method. It was generally agreed that the EQ-5D was not appropriate in evaluations of treatments for children, or treatment for hearing, vision or mental health conditions, and that more guidance should be included in the Methods Guide on this.

- **How to identify the most appropriate measures (EQ5D or other):** Participants did not conclude how utility values should be identified. It was stated that ideally these should be identified in the same way as clinical data, but that this was not considered to be practical or do-able. There was no consensus on how systematic/exhaustive searches for utility values should be. However, there was overall agreement that searches should be transparent, explicit and reproducible.

  It was generally agreed that trial-based utility values should not be used in isolation, and that evidence providers should attempt to validate these values with values identified in the literature, explore any differences, and address issues related to extrapolation and generalisability.

  Participants agreed that no established quality criteria currently exist to choose most appropriate utility values. Participants agreed that synthesising or pooling utility values from different sources was not appropriate, and that the most important criterion for selecting a utility value is the relevance to the health state being modelled. It was generally agreed that utility values that are selected are fully justified, and that any uncertainty should be explored in sensitivity analysis.

- **Mapping**: Participants agreed that the mapping function chosen should be fully described, its choice justified, and it should be adequately demonstrated how well the function fits the data. Sensitivity analyses should be presented if there are several algorithms in the literature but no clear preference over which algorithm should be used, and to explore any uncertainties in the mapping algorithm.

During the workshop discussions it was frequently suggested that the utility values for a particular appraisal could be determined during the scoping stage between the evidence provider and NICE. Furthermore, it was suggested that NICE could set up a publicly available archive of previously used utility values, exceptions to the EQ-5D accepted by the Appraisal Committee, and what alternatives, including specific mapping algorithms, have been accepted by the Appraisal Committee in which circumstances.

The following 3 additional topics were also addressed:

- **EQ-5D 5L:** Participants agreed that NICE could not currently use the EQ-5D-5L as reference case because it is yet not fully evaluated. However, participants felt that NICE needs to make statements about the usefulness of measuring quality of life in parallel with both the 3L and the 5L version. This was because participants were concerned about any problems with the transition from the 3L to the 5L version bearing in mind the time it would take to collect EQ-5D data before a NICE appraisal.

- **Adjustments to utility values:** Participants agreed that in some circumstances adjustment, for example for age or co-morbidities, to utility values could be required in order to provide unbiased estimates of health effects. However, it was stated that the appropriate method by which to undertake such adjustment, and the impact of it, will vary by the context and the technology. Also, because of the current ongoing methodological debate about adjustments, participants felt that the Methods Guide should not be too prescriptive in respect of utility adjustments, but expressed the importance of a pragmatic approach tailored to the data used in each appraisal.

- **Quality of life benefits for people other than patients and carers:** Although the reference case stipulates 'all health effects on individuals' the text in section 5.2.7 mentions *'patients or, when relevant, other people (principally carers)'*. There was no consensus at the workshop about the appropriateness of this wording. Some participants thought that only health effects on patients should be taken into account, some thought that not

only carers but also significant others (e.g. children, grandparents, parents, foster parents, brothers, sisters, companions, dependents), and some thought that section 5.2.7 was appropriate, as carers are performing a different role to that of significant others, i.e. one which in some cases would have to be provided by the NHS or Social services. Participants expressed concern that including significant others would favour conditions affecting people of child bearing age who are more likely to have dependents. Participants, however, agreed that if any health benefits beyond patients are included, it has to be based on empirical data, rather than claims only. Participants also felt that there were methodological problems to address with aligning tools to measure carer quality of life with the EQ-5D. As far as non-health quality of life benefits are concerned, participants agreed that this depends on the final decision on the perspective and needs to be consistent across all individuals whose health effects are included.

There was much discussion about of the **DSU's technical support documents** (TSDs) and the need to clarify the role of these in relation to the Methods Guide, whether they should be referenced or the technical advice embedded in the Methods Guide. Some Participants raised the concern that these documents have not been consulted upon. The specific TSDs mentioned were

- TSD 9 ('The identification, review and synthesis of health state utility values from the literature')
- TSD 11 ('Alternatives to EQ-5D for generating health state utility values')
- TSD 12 ('The use of health state utility values in decision models')

2 Questions posed to the workshop participants

1. Should the EQ-5D be the reference case measure across all patient populations (e.g. adults, children) and diseases? Are there any exceptions in which an alternative measure (e.g. the SF-6D or the HUI-3) should be used? If such exceptions arise, which alternative measures should be recommended? How should NICE's Appraisal Committee deal with appraisals whereby the available reference case utility values are not considered plausible?

2. How, and when, should NICE adopt the new EQ-5D-5L tariff? Should NICE follow the lead of the Euroqol group or should it set its own agenda?

3. When, if ever, should mapping (cross-walking) be preferred over the direct valuation of health (e.g. using the EQ-5D)? Can a consistent set of criteria be set out to define such circumstances? What are these criteria? Should such analyses be presented as the base case analysis or as a secondary analysis? How can the Appraisal Committee ensure that the mapping is reasonable?

4. How should potentially relevant utility values be identified? Is a systematic review of utility studies necessary? How should appropriate utility values be selected? How should values be synthesised across studies such that the uncertainty is adequately reflected)?

5. Should models reflect changing utility over time (for example, between disease progression and death), and if so how? Should utility values for health states be adjusted for age and/or sex when incorporated into economic models?

6. How should health effects on people other than the recipient of the intervention (e.g. parents, carers) be defined, measured and valued within technology appraisals?

3 Summary of the workshop discussions

3.1 The EQ-5D as the reference case

The current Methods Guide states in section 5.4.1 that '*The EQ-5D is the preferred measure of HRQL in adults. The methods to elicit EQ-5D utility values should be fully described. When EQ-5D data are not available or are inappropriate for the condition or effects of treatment, the valuation methods should be fully described and comparable to those used for the EQ-5D.'* It also states a preference for the *EQ-5D* in section 5.4.4. Section 5.4.9 and 5.4.10 explain in more detail what is required if the EQ-5D is not considered appropriate (see full text in Appendix 1).

3.1.1 *Should the EQ-5D be the reference case measure across all patient populations (e.g. adults, children) and diseases?*

In general, participants felt that the EQ-5D is an appropriate reference case because it is a standardized tool, based on rigorous research, the most widely used measure, and because it works in most cases. However, views differed on the extent to which in exceptional cases other measures can be used instead or alongside the EQ-5D. Some participants were of the view that no exceptions should be permitted because of the need for comparability and consistency. However, the majority of participants agreed that exceptions to the reference case would be appropriate if there is well substantiated evidence that the EQ-5D is not suitable for the particular patient population or disease.

3.1.2 *What could be exceptional circumstances?*

In general, it was felt that the Methods Guide needs to be very explicit about which exceptions are allowed and to guide the manufacturer on how to demonstrate which situations may be exceptions.

Some participants felt that it would only be appropriate to use another measure if evidence from literature (academic publications, rather than unpublished analyses presented by the manufacturer) demonstrates that the

EQ-5D should not be used in specific circumstances. Some participants suggested that the EQ-5D should always be used and if a manufacturer wanted to present disease specific or other measures, these should only be included in sensitivity analysis. This is because the EQ-5D was considered to be a reference standard – and that it is not possible to assess alternative utility values if there is nothing to compare them with.

Some participants thought that the EQ-5D is generally not appropriate for children and that the Methods Guide should provide more guidance. Participants were aware of the children's version of EQ-5D descriptions (EQ-5DY) but that this was not valid for children under 5 years of age. Participants felt that there is the need for research on whether valuations obtained from adults are applicable to children. Some participants questioned whether children value health differently to adults, but then generally agreed that it is appropriate for the valuation to be carried out by adults, as these are the tax-paying general population.

Participants stated that two important areas of health-related quality of life which are not captured with the EQ-5D were hearing and vision because loss of these two senses may matter to the patient in more ways than affecting mobility, usual activity, and anxiety and depression. It was suggested that the HUI may be better in these circumstances, but by allowing other measures could give advantages to specific diseases (if the instruments are chosen that detect the largest differences). It would therefore be important to use EQ-5D alongside disease specific questionnaire.

Other specific examples where participants felt the EQ-5D might not work were:

- Mental health – potential problems with patients self-reporting
- Co-morbidities - EQ5D not considered valid as no able pick up differences
- Quality of life in diagnostics

- Ultra orphan diseases – lack of data, epidemiology very poor.

3.1.3 *Which alternative measures should be recommended?*

Participants agreed that disease specific measures should be used when EQ-5D does not capture all relevant dimensions, but again, evidence that this is so would need to be provided.

Participants also felt that if other measures are presented alongside the EQ-5D then the valuation method of this alternative descriptive system needed to be comparable to how it is done with the EQ-5D, i.e. using TTO. Also, any direct valuation of health states could also be used using TTO, but in this case the disease is known and some diseases carry more emotional weight than others, which may change the way patients rank or value health. With generic descriptions this does not happen.

Participants noted that the EQ5D valuation maybe out of date (1996) with societal preferences having moved on. Also, some participants felt that it was important to mention that EQ-5D is not aimed to measure functioning unlike some disease specific measure and that it this needs to be borne in mind to avoid confusing the sensitivity of the EQ-5D in detecting functional impairments and in detecting changes in HRQoL.

Participants were aware that for vision and hearing ‘bolt ons’ to the EQ-5D may be an option but that more research is needed on this methodology and the consequences of its use.

Some participants suggested a list on NICE’s website with the exceptions accepted by the Appraisal Committee – this list could be updated as more evidence becomes available. Others suggested a discussion about alternatives to be included at the scoping stage.

Participants felt that the DSU’s Technical Support Document (TSD) 11 (‘Alternatives to EQ-5D for generating health state utility values’) provides a good basis for this and should be referenced in the methods guide.

3.1.4 *How should NICE's Appraisal Committee deal with appraisals whereby the available reference case utility values are not considered plausible?*

If the available reference case utility values are not considered plausible, participants suggested that the Committee need to use common sense in considering utility values obtained, taking into account the accumulated experience of previous appraisals. Also, exploring the importance of the utility values for the cost effectiveness through sensitivity analysis was considered appropriate.

Participants felt that if significant difference between values obtained from alternative or condition-specific measures and from EQ-5D were found that more explorations of the reasons are needed. This may involve requesting more information from the manufacturer. It may also be necessary for the Committee use judgment and deliberation, but that this needs to be explained well.

3.2 How, and when, should NICE adopt the new EQ-5D-5L tariff? Should NICE follow the lead of the EuroQol group or should it set its own agenda?

Participants agreed thought that NICE could not recommend the EQ-5D-5L as the reference case at this moment in time, but expressed the view that NICE could not ignore the EQ-5D-5L, and that some guidance was required. Specifically, some groups thought that if NICE had specific views on how the valuation of the EQ-5D-5L should be undertaken, it should request such research proactively. Concerns were raised that NICE's view on the EQ-5D-5L could have implications for the continued development of the EQ-5D-5L. Many groups confirmed that a signal of support from NICE could aid its development internationally, and support the development of more evidence.

Participants suggested that before NICE can consider the incorporation of the *EQ-5D-5L* as the reference case, evidence needs to be available

- about the sensitivity of the EQ-5D-5L,
- that people can appropriately differentiate between the levels,
- to map from EQ-5D-5L to the EQ-5D-3L values, and
- that addresses the issue that the 55555 state is unlikely to be equal to the 33333 due to changes in the method of valuing states worse than dead.

Many groups discussed the time lags between the EQ-5D-5L validation by the EuroQol group, generation of data and the adoption by NICE, and agreed that industry needs guidance now for products that may be appraised in several years. There were concerns that 'parallel running' of EQ-5D-3L and EQ-5D-5L in trials could lead to gaming. However some groups thought that data produced in parallel would be informative.

Participants agreed that any EQ-5D-5L generated should be presented in a sensitivity/secondary analysis, with EQ-5D-3L always being presented as base case.

There was some discussion about the valuation methods, which EuroQol may use, but there was no clear consensus about this. Some groups voiced concerns about the TTO method, and others thought that discrete choice evaluation was a promising methodology for this new valuation.

3.3 Mapping

3.3.1 *When, if ever, should mapping (cross-walking) be preferred over the direct valuation of health (e.g. using the EQ-5D)? Should such analyses be presented as the base case analysis or as a secondary analysis?*

The current Methods Guide states in section 5.4.6 that '…. *When EQ-5D data are not available, methods can be used to estimate EQ-5D utility data by mapping (also known as 'cross-walking') EQ-5D utility data from other HRQL measures included in the relevant clinical trial(s). This can be done if an adequate mapping function can be demonstrated and validated. Mapping*

should be based on empirical data and the statistical properties of the mapping function should be clearly described.'

In general, workshop participants considered that mapping should only be undertaken in exceptional circumstances such as when:

- EQ-5D is not suitable
- The study from which the health-related quality of life measures were derived is poorly designed or has a very small population (e.g. for rare diseases)
- A specific quality of life measure is collected in the trial at time points which cannot reasonably inform the model

Some participants noted that it would be preferable to use an existing mapping algorithm (if it has been appropriately validated and is still up to date). However, participants also expressed concern that often manufacturers use algorithms which have been previously considered and accepted by Committee, even if they are now largely out of date, or contain an error. In light of this, participants agreed that the mapping function chosen should be fully described, its choice justified, and it should be adequately demonstrated how well the function fits the data. Any other available mapping functions should also be included as secondary analyses. Uncertainty around the mapping function used should also be clearly described and tested in sensitivity analyses.

Some participants debated whether mapping to the utility values or to the dimensions (response mapping) is more appropriate. It was noted that response mapping is not widely undertaken in the UK at present; however it does have advantages in being able to include a comparison of patients across diverse instruments and in having flexibility in the degree of precision desired.

3.3.2 *Can a consistent set of criteria be set out to define such circumstances? What are these criteria? How can the Appraisal Committee ensure that the mapping is reasonable?*

Participants considered that if mapping is required, there are likely to be significant variations in the methodology used unless explicit instructions are provided to the manufacturer. Participants stated that the current Methods Guide does not adequately describe when the EQ-5D is not an appropriate measure and what criteria determine which preferred measure should be presented instead. They suggested that a hyperlink to the DSU's TSD 12 ('The use of health state utility values in decision models') should be embedded in the methods guide, and a summary of key points from the TSD should also be presented, to help readers understand how to determine which method to use if EQ-5D is not available/not appropriate.

Some participants suggested that the modelling developers should advise NICE during the scoping process whether utility values were directly collected in the key trials, or whether a mapping function will be used. NICE should then advise on the most appropriate approach to derive utility values. However, concern was expressed that some Assessment Teams do not have mapping specialists in their teams and therefore it would be challenging for them to determine the most appropriate mapping function to use.

Participants suggested that advisory meetings could be held to determine which mapping functions are most appropriate for each therapeutic area. Additionally, a systematic review of all published mapping functions should be undertaken. One participant cited work from Danny Frybach (using a US database of underlying health state measures) and suggested that NICE should review it and use it to inform which mapping algorithms should be considered if EQ-5D is not available.

Participants also highlighted the benefit of producing an archive (or publically available database) which contains all previous mapping functions used in technology appraisals, alongside a list of criteria to determine the most appropriate function to use for each therapeutic area. This would also ensure

a systematic collection of the strengths and weaknesses of previous mapping methods used, and assist the Committee with the decision about the appropriateness of the mapping function used in an appraisal.

Participants concluded that more explicit instructions should be included in the methods guide and in the NICE submission template to help justify the choice of mapping function and adequately explore and describe any uncertainties. In particular, it is important that the sample used to derive the health-related quality of life measures adequately matches the patient population under consideration in the appraisal (ideally all health states should come from the source data) and cover all disease states. In addition, any mapping algorithm should be externally validated (preferably on a separate patient sample) and results from model fit tests be provided.

3.4 How should potentially relevant utility values be identified?

The current Methods Guide states in section 5.4.11 that '*When health-related utility values have been obtained from the literature, the methods of identification of the data should be systematic and transparent. The justification for choosing a particular data set should be clearly explained. Health-related utility data that do not meet the criteria for the reference case should be accompanied by a carefully detailed account of the methods used to generate the data and a consideration of how these methods may affect the values. When more than one plausible set of h health-related utility data are available, a sensitivity analysis should be undertaken.'*

Participants discussed a number of different ways to identify utility values:

- **Searches**: There was no overall consensus on how systematic/exhaustive searches should be. Participants felt that given time constraints involved in the appraisal process, systematic literature search methods may not always be feasible. However, there was overall agreement that searches should be transparent, explicit and reproducible e.g. for the ERG/Assessment Group to re-run the search.

Participants could not find a consensus about how prescriptive NICE should be in defining a systematic search strategy and a selection process and whether evidence providers should be offered more explicit guidance on how systematic reviews should be conducted e.g. from Decision Support Unit technical support documents.

- **Clinical trials**: Often the manufacturer will decide a priori to collect utility scores within the clinical trial that is used as part of their submission. There was general agreement that such data would probably be the most useful in capturing HRQoL impact for the relevant population in the submission. However, there were issues about extrapolating utility values measured from a relatively short time period (within a clinical trial) over a longer term horizon, as required in an economic model. It was generally agreed that trial-based utility values should not be used in isolation and that evidence providers should attempt to validate these values with utility values identified in the literature and explore any significant differences. There was also concern on the generalisability of the patient population in the clinical trial to the UK NHS setting.

- **Electronic Databases**: Participancts suggested that existing databases e.g. PROMS may be used. However, there were concerns about how comprehensive such databases were in terms of the disease areas covered. Two concerns were raised against using PROMs databases: (1) that patient heterogeneity was masked due to the large patient numbers, and (2) that the narrow confidence intervals were not representing the true uncertainty around the estimates.

- **Expert elicitation:** Participants generally agreed that this should be done only if utility values are not identified in a literature search or collected within a trial.

- **Previously published Technology Appraisals**: As there are a growing number of appraisals which include utility values as part of the

economic model, one participant suggested that a database of utility values used in published appraisals should be set up by NICE.

- **Previously published systematic reviews:** Participants agreed that if there has been a recent, well-conducted systematic literature review already conducted either in a previous appraisal or journal article, it would be reasonable to identify relevant utility values from these sources. This would avoid duplication of effort especially given time constraints involved in the appraisal process. However, it was agreed that older reviews (e.g. > 5 years?) may be out-of-date and that using poorer quality reviews may lead to replication of errors. Therefore, some quality assessment of previously published reviews may be necessary.

3.4.1 *Quality criteria for the selection of utility values*

Participants agreed that no established quality criteria currently exists (e.g. quality checklists) in order to choose most appropriate utility values. There was some discussion of what such quality criteria should be, e.g. sample size, missing data, country, type of instrument, generalisability to the patient population in the UK NHS and internal validity.

However, relevance of the utility values to the health states that are being modelled was seen as the most important issue when selecting utility values. There was however, overall agreement that cherry-picking utility values (selection bias), with no explicit or transparent method of identification, should be avoided. It was generally agreed that utility values that are selected are fully justified and that any uncertainty should be explored (in sensitivity analysis)

3.4.2 *How should values be synthesized across studies such that the uncertainty is adequately reflected?*

Participants discussed if formal methods of data synthesis including pooled utility values should be used when there are sufficient number of homogenous utility values available (e.g. from same patient population and using the same

instrument). Some form of meta-regression may also be useful to explore any causes of variation between utility values. However, it was agreed that this will not always be feasible if the degree of heterogeneity is too high. It was suggested by some workshop participants that identifying the most relevant single utility value is more crucial than attempting to pool or synthesise multiple utility values when a large number of potentially relevant values are identified in the literature.

Participants acknowledged that ongoing MRC-funded research is in progress to explore evidence synthesis methods applied to utility values but does not report until the end of 2012.

It was suggested that it was more important to explore any uncertainty in the utility difference between relevant health states rather than the uncertainty in any baseline utility values.

Participants generally agreed that current methods to deal with uncertainty around utility values may be appropriate e.g. using alternative utility values if available. If alternative utility values are unavailable then threshold analyses may be appropriate i.e. varying the utility values between plausible ranges to explore how the ICERs are affected.

There was reference to the DSU's TSD 9 ('The identification, review and synthesis of health state utility values from the literature') but participants were uncertain on its role as guidance documents in relation to the NICE methods guide.

3.5 Should models reflect changing utility over time (for example, between disease progression and death), and if so how? Should utility values for health states be adjusted for age and/or sex when incorporated into economic models?

The current Methods Guide does not make reference in section 5.4 to any adjustments that may be necessary when incorporating utility values into economic models.

This topic was added after the development of the briefing paper for this workshop and therefore the briefing paper did not cover the topic of potential adjustments to utility values. This meant that the responses from the workshop participants to this question may not have been as focused as otherwise possible.

Participants acknowledged that there is ongoing debate in the health economics community about appropriateness of adjusting utility values. The overall consensus was that a pragmatic approach should be taken and the Methods Guide should incorporate flexibility to allow the most appropriate approach tailored to the individual appraisal.

Participants focussed their discussions on adjustments for varying utility over time, age and sex. Generally, the majority of participants expressed the view that it was appropriate to vary utility over time in the modelling of health benefits and that this is already the current practice within NICE appraisals. However, some participants felt that such an adjustment was rare and should not normally happen in NICE appraisals.

It was commonly accepted that age adjustment is in fact a proxy to adjust for the average increased comorbidity and decline of function (e.g. with respect to eyesight or hearing), which generally occurs as people age. However, some delegates explained that often it is difficult to separate the effect of disease progression from disease-unrelated comorbidities, and that the impact of this can vary according to the disease area. Also, participants discussed that multiple related comorbidities may have a lesser impact on quality of life than two completely unrelated comorbidities.

Participants expressed the opinion that ideally models should contain data from a representative mix of people of all ages and disease severity as appropriate. Where this is available, directly observed utility values should be used instead of age-adjusted or comorbidity-adjusted values, but it was accepted that in most cases such data are not available. Most participants expressed the view that in this situation, the utility values should be adjusted for age in order to describe plausible health gains. Moreover, some participants expressed the view that face validity of models could be compromised if adjustment is not performed, particularly when modelling over an extensive period of time.

Some participants noted that the Methods Guide currently stipulates that future costs that are considered to be unrelated to the condition or technology of interest should be excluded from the evaluation. These participants therefore questioned whether the adjustment of utility values for future unrelated comorbidities might imply a different perspective for health effects compared to costs. However, other participants suggested that the inclusion of changes in health unrelated to the condition or technology of interest was required in order to estimate overall health gains from the technology of interest and that this did not constitute an inconsistency in terms of perspective.

Some participants were concerned that adjustment for age could be interpreted as indirect age discrimination. Other participants expressed the view that age-adjustment was not discriminatory because it captured the average increase in comorbidity as people get older, and that adjusting was the correct methodological approach to capture benefit over time.

Participants also discussed adjustment for sex and stated that the difference in life expectancy between men and women could affect accumulated life years, and subsequently any discounting of QALYs could therefore have a differential impact on women and men. Furthermore there could be a difference in the natural history of illness between men and women and this could also cause a differential effect. To mitigate this, participants felt that gender-specific utility data would ideally be used, however conceded that data

on both condition-specific and health-specific utility values were unlikely to be available for most appraisals.

In summary, participants from more than one group expressed the importance of a pragmatic approach to this problem and that in some circumstances adjustment could be required in order to provide unbiased estimates of health effects. Participants felt that the Methods Guide should not be too prescriptive in respect of utility adjustment. It was stated that the appropriate method by which to undertake utility adjustment, and the impact of any adjustment, will vary by the context and the technology. There was also discussion about the unresolved issue of the implementation of various methods, including methods that are multiplicative, additive, mixed and non-linear. Therefore, participants felt that the best option would be to take an approach that does not restrict the appraisal to a particular method of adjustment. Many participants stated the need for presentation of ICERs including both adjusted and unadjusted utility values in the economic model. This would illustrate the impact of the adjustment on the estimated health effects, and would be valuable in cases where the Appraisal Committee deemed that the method of adjustment used was inappropriate.

Some delegates suggested that the Methods Guide could make reference to the DSU's TSD 12 ('The use of health state utility values in decision models') that provides more information on appropriate methods. However, other participants noted that the TSDs are not currently put out for public consultation, unlike the Methods Guide.

3.6 Should the impact on significant others be broadened out to include other members of the family who are not directly involved in care

The current Methods Guide states that '*all health effects on individuals' should be taken into account* (table 5.1; page 30) and section 5.2.7 specifies that '*the perspective on outcomes should be all direct health effects, whether for patients or, when relevant, other people (principally carers).*'

Participants at the workshop generally felt that it was appropriate to consider benefits both to carers and to significant others, but had a number of concerns which may be difficult to overcome.

Participants preferred that benefits to significant others should be included as part of the deliberative process, rather than including these benefits formally in the economic model, and that if it is done it should be based on empirical evidence. This was because of the limitations to current methodology for including benefits to carers and significant others in economic evaluations and difficulties in data collection. In general, participants were unclear about the likely effect on cost effectiveness of including such data and whether the benefits of including such data would be outweighed by the effort to collect it.

Participants raised concerns about where the boundary of significant others would be drawn (e.g. children, grandparents, parents, foster parents, siblings, co-habitants, dependents), and that without clear boundaries this could lead to great inconsistencies if some evidence provides include more ‘beneficiaries’ than others.

Participants considered that including health benefits to significant others raised equalities issues in that such an approach could favour some conditions over others such as those affecting people of child bearing age who are more likely to have a number of dependents. It was noted by one PCT attendee that factors such as family status were not considered in individual funding requests for equalities reasons.

These concerns were felt to be particularly important in the context of the introduction of value based pricing where the price of a product could be influenced by these factors, thereby potentially incentivising evidence providers to identify and incorporate such benefits. There was one suggestion that it could be left for the manufacturers/sponsors to submit such evidence if they wished. Other participants, however, felt that with the introduction of value based pricing there was a need for consistency and more specific guidance from NICE.

Some participants suggested that if benefits to carers and other significant others were included, this would affect opportunity cost and would affect the cost-effectiveness threshold.

Participants noted that if evidence was available for health effects on other people but the patient that could be included in an economic evaluation then this may be most appropriately submitted as a sensitivity analysis to the reference case.

There was no consensus about the appropriateness of the wording in the current methods guide. Specific points raised were:

- the current reference case allowing the inclusion of carer benefits in economic modelling is not appropriate, and that the economic modelling should focus on the benefits to the patient.
- the current methods guide is inconsistent and that if NICE accepts carer benefits then for consistency, it should be all significant others and not limited to those of carers.
- the current methods guide is appropriate, as carers are performing a different role to that of the non-caring wider family, one which in some cases would have to be provided by the NHS or Social services if it didn't exist.

3.6.1 *Extend benefits to non-health-related quality of life*

Participants also discussed if only health effects or non-health-related quality of life effects should be taken into account, and noted that this question was related to the decision about the most appropriate perspective, and whether the perspective should remain that of the NHS or be extended to a wide societal perspective. Participants agreed that if non-health-related quality of life was included for carers and significant others, then this should also be done for patients. Likewise if only health-related benefits were considered for patients, then this should also be done for carers and significant others.

3.6.2 *Aggregation of impacts on carers and significant others*

Participants also discussed how any impacts on carers and significant others should be aggregated and agreed that there would be challenges in collecting and analysing data from carers and significant others. For example to enable collection of data from carers or significant others in a trial, such people would need to formally consent to, and be enrolled in, the trial. This would add additional administration and cost to trials.

If only health benefits were included, some concern was raised about whether the EQ-5D would be sufficiently sensitive to identify the impact on carers and significant others. Further it was noted that currently care-specific measures include a more general focus on quality of life rather than health related quality of life and are not anchored to 0 and 1 in the same way as the EQ-5D. Participants considered that it would not be possible to aggregate different measures for example EQ-5D for patients and CarerQOL for carers. Participants also questioned whether it is appropriate to assume an equal weight for benefits to carers and significant others compared to benefits to patients. For all these reasons, before data for carers or significant others were to be formally considered, there is a need for methodological research for this to be done appropriately.

Participants felt that these methodological issues meant that currently it would not be appropriate for NICE to make quantification of benefits to carers and significant others a requirement. Instead, the consensus was that that the Appraisal Committee should deliberate these benefits.

4 Key issues for consideration by Working party

1. Should the Methods Guide be more descriptive than in the current sections 5.4.9 and 10 about the circumstances in which utility measures other than EQ-5D are acceptable?

2. If so,

    a) What are these circumstances?

b) What supporting evidence needs to be provided?

3. Should decisions about alternative utility measures be made at the scoping stage of an appraisal?

4. Should decisions about alternative utility measures be made be based on ‘case law’ developed through previous appraisals?

5. Should the EQ-5D always be presented alongside any alternative measures?

6. Should NICE encourage the parallel use of EQ-5D 5L?

7. If so which data should be used for decision making in an appraisal where both sets of data are available?

8. Should more information on mapping than in the current section 5.4.6 be included in the Methods Guide?

9. Should the Methods Guide be more explicit about when utility values need adjusting, for example for age, and if so, how to carry out such adjustments?

10. Should in the definition of who benefits (‘all health effects on individuals’) be changed to be more specific about who those individuals are?

11. Should there be a difference between carers and significant others?

12. Should the only health benefits for people other than patients be included?

13. Should the DSU’s TSDs be included/ embedded/ referenced in the Methods Guide?

5 Authors

Prepared by Elisabeth George on the basis of workshop feedback from Moni Choudhury, Matthew Dyer, Rita Faria, Zoe Garrett, Susan Griffin, Martin

Hoyle, Pall Jonsson, Brendan Mulhern, Fiona Rinaldi, Marta Soares, Jon Tosh and Helen Tucker, whose contributions are gratefully acknowledged.

January 2012

6 Appendix 1

Extract from the Current Methods Guide

Section 5.4 Measuring and valuing health effects

5.4.1 For cost-effectiveness analysis, the value of health effects should be expressed in terms of QALYs for the appropriate time horizon. For the reference case, the measurement of changes in HRQL should be reported directly from patients and the value of changes in patients' HRQL (that is, utilities) should be based on public preferences using a choice-based method. The EQ-5D is the preferred measure of HRQL in adults. The methods to elicit EQ-5D utility values should be fully described. When EQ-5D data are not available or are inappropriate for the condition or effects of treatment, the valuation methods should be fully described and comparable to those used for the EQ-5D. Data collected using condition-specific, preference-based measures may be presented in separate analyses. The use of utility estimates from published literature must be supported by evidence that demonstrates that they have been identified and selected systematically.

5.4.2 The QALY is a measure of a person's length of life weighted by a valuation of their HRQL over that period. The HRQL 'weighting' usually comprises two elements: the description of changes in HRQL itself and a valuation of that description of HRQL. Information on changes in HRQL as a result of treatment should be reported directly by patients (and directly by carers when the impact of treatment on the carer's health is also important). The valuation of changes in HRQL reported by patients should be based on public preferences elicited using a choice-based method in a representative sample of the UK population.

5.4.3 When it is not possible to obtain information on changes in patients' HRQL directly from patients, then data should be obtained from their carer (not from healthcare professionals). The valuation of changes in

HRQL measured in patients (or carers) should be based on a valuation of public preferences from a representative sample of the UK population.

5.4.4 To quantify the effects of technologies on HRQL, the EQ-5D (a standardised and validated generic instrument) is preferred. Different classification systems produce different utility values; therefore, results from the use of different systems cannot always be compared. Given the comparative nature of the Institute's work and the need for consistency across appraisals, a single classification system, the EQ-5D, is preferred for the measurement and valuation of HRQL.

5.4.5 The EQ-5D is a widely used measure of HRQL and has been validated in many different patient populations. The EQ-5D comprises five dimensions of health: mobility, ability to self-care, ability to undertake usual activities, pain and discomfort, and anxiety and depression. The system has been designed so that people can describe their own HRQL using a standardised descriptive system. A set of preference values elicited from a large UK population study using a choice-based method of valuation (the time trade-off method) is available for the EQ-5D classification system. This set of values can be applied to people's self-reported descriptions of their HRQL to generate health-related utility values.

5.4.6 Data using the EQ-5D instrument may not always be available. When EQ-5D data are not available, methods can be used to estimate EQ-5D utility data by mapping (also known as 'cross-walking') EQ-5D utility data from other HRQL measures included in the relevant clinical trial(s). This can be done if an adequate mapping function can be demonstrated and validated. Mapping should be based on empirical data and the statistical properties of the mapping function should be clearly described.

5.4.7 When EQ-5D utility data are not available, direct valuations of descriptions of health states based on standardised and validated HRQL measures included in the relevant clinical trial(s) may be submitted. In these cases, the valuation of descriptions should use the time trade-off

method in a representative sample of the UK population, with ‘full health’ as the upper anchor, to retain methodological consistency with the methods used to value the EQ-5D.

5.4.8 Data that have been collected directly in relevant clinical trials using condition-specific, preference-based measures should be presented in a separate economic analysis.

5.4.9 The EQ-5D may not be an appropriate measure of health-related utility in all circumstances. If the EQ-5D is considered inappropriate, empirical evidence should be provided on why the properties of the EQ-5D are not suitable for the particular patient population. These properties may include the content validity, construct validity, responsiveness and reliability of EQ-5D. When an alternative measure is preferred, those submitting analysis should provide reasons, supported by empirical data on the properties of the instrument used. They should also indicate any evidence that will help the Committee understand to what extent their choice of instrument has impacted on the valuation of the QALYs gained. If direct valuations of descriptions of health states based on HRQL measures other than the EQ-5D are used, the valuation methods must be comparable to those used for the EQ-5D (see section 5.4.5).

5.4.10 It is recognised that the current version of the EQ-5D has not been designed for use in children. When necessary, consideration should be given to alternative standardised and validated preference-based measures of HRQL, such as the Health Utility Index 2 (HUI 2), that have been designed specifically for use in children.

5.4.11 When health-related utility values have been obtained from the literature, the methods of identification of the data should be systematic and transparent. The justification for choosing a particular data set should be clearly explained. Health-related utility data that do not meet the criteria for the reference case should be accompanied by a carefully detailed account of the methods used to generate the data and a consideration of how these methods may affect the values. When more

than one plausible set of h health-related utility data are available, a sensitivity analysis should be undertaken.