

NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Report to the Methods Review Working Party

Key issues arising from workshop on QALY weighting

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

Participants at the workshop addressed a number of questions raised by the briefing paper in five groups facilitated by representatives of the NICE decision support unit.

The workshop discussions addressed four distinct topics:

- Appropriateness of QALY weighting
- Identifying the type of criteria which should be weighted
- Determining weights for specific criteria and incorporating them into the appraisal process
- Opportunity cost issues arising from QALY weighting

Most workshop participants considered that QALY weighting should only be conducted for exceptional cases (5-10% of appraisals), and that it would

serve to achieve greater transparency around decision-making and improved consistency between appraisals. Participants suggested that NICE should look to the Government's Value Based Pricing proposal and the NHS Operating Framework when determining which criteria would be most appropriate for QALY weighting. It was also suggested that the ongoing work by the University of Sheffield to derive QALY weights for these criteria should also be considered.

There was little support among participants for making explicit changes to the current threshold in order to allow for formal consideration of QALY weighting in the cost effectiveness framework.

2 Questions posed to the workshop participants

1. Do you think that QALY weighting is reasonable for NICE to do?
Should QALYs be weighted at all in NICE appraisals? If not, why?
2. Which criteria are those that should be considered most appropriate for QALY weighting? (For example, end of life, severity, children, size of benefit, unmet need, improvement and innovation, rarity, others). How should these criteria be defined?
3. Who should decide on the QALY weights? How should they be elicited?
4. How should QALY weights be incorporated into the assessment of new technologies? Should they be incorporated analytically into the economic analysis or should they be dealt with as part of the deliberative process?
5. If QALY weighting is formally incorporated into the cost-effectiveness framework, how should NICE deal with the opportunity cost issues arising from QALY weighting? Should NICE:
 - I. Adjust the threshold?
 - II. Apply the weights only in very rare situations such that the threshold would be largely unaffected?

III. Apply weights symmetrically within appraisals such that there are both positive and negative weights?

IV. Other?

How should NICE consider the issue of opportunity cost if QALY weights are reflected in a less formal manner?

3 Summary of the workshop discussions

3.1 *Appropriateness of QALY weighting*

Workshop participants acknowledged that the current deliberative process, which allows for implicit weighting, is generally sufficient and allows for flexibility in decision-making. However, some participants expressed the view that deliberation can lead to inconsistent decision making between appraisals and that QALY weighting may help achieve greater consistency and transparency.

Overall most workshop participants considered that QALY weighting should only be conducted for exceptional cases (5-10% of appraisals). In such instances, there should be a strong argument to justify any deviations from the base case (that is, any deviation from a QALY weighting of 1). Some participants also expressed the view that weights should also be applied to QALYs of displaced activity in the NHS.

3.2 *Identifying the type of criteria which should be weighted*

There was a general consensus that the purpose of QALY weighting is to incorporate elements about the technology that are not currently included in the QALY. Any criteria chosen for additional weighting should avoid double counting what is already included in the QALY. Some participants were also concerned that most criteria are correlated and therefore cannot be considered separately.

The participants considered that there should be some kind of ethical basis or justification behind any criteria which receive additional QALY weighting, and that criteria should be very stringently defined so that, in practice, additional

QALY weighting would only be applied in exceptional cases rather than in the majority of technology appraisals.

Participants noted it would be easier to weight some criteria (for example certain populations such as children) more than others (for example innovation). It was also noted that if additional QALY weighting were accepted for Technology Appraisals, this would inevitably have an impact on all of the other guidance producing programmes.

Which criteria should be weighted?

1. **Severity of disease.** Participants noted that severe diseases (such as cancer) are likely to be preferentially weighted if QALY weighting is introduced. There was uncertainty surrounding how severity of disease should be defined.
2. **Children.** It was noted that the public might attach more weight to children because they have a longer life expectancy; however this represents double counting as lifetime gains are already included in the QALY. Participants also noted however that even children who have a short life expectancy are still likely to have more weight attached to them by the public.
3. **Age.** Participants discussed the 'fair innings' argument that a lower QALY weighting could be applied to elderly people, although it was queried as to whether QALY weighing should also be applied to middle aged adults. Delegates noted that the majority of serious illnesses arise in people older than 55 years so they considered whether a lower weight could be applied to people over 55 years of age. Overall, there was limited support for a reduced QALY weighting in the middle aged or elderly adults.
4. **Size of benefit.** Most participants did not consider it appropriate to weight size of benefit as it is already included in the QALY gain. It was noted however the Government's Value Based Pricing proposal will consider magnitude of therapeutic improvement, and therefore participants considered that it would be important for NICE to ensure

that the Committee considers criteria which are consistent with the Government's proposal.

5. **Personal responsibility.** Most delegates did not consider it appropriate to negatively weight conditions which are associated with lifestyle choices, for example smoking, alcohol or drug misuse or obesity. There was recognition that it is difficult to prove a causal link between some activities and diseases.
6. **Rarity.** None of the participants considered it appropriate to give additional weight to orphan or ultra orphan diseases. Although it was noted that the objective of weighting rare disease was to incentivise drug development, participants did not think it appropriate to do this by QALY weighting.
7. **Innovation.** None of the participants considered it appropriate to weight innovation because they considered that a 'step change' in the management of a disease should already be captured (to some extent) in the QALY calculation. The participants noted that some of the benefits of innovation may be captured in other criteria such as unmet need. They also noted that often the innovative nature of a technology has no impact on the patient beyond what is already measured in the QALY. Although participants did not consider it appropriate to weight innovation, they did consider it useful for innovation to be taken into consideration in the Committee's deliberations.
8. **Unmet need.** None of the participants considered it appropriate to weight unmet need. It was noted however the Government's Value Based Pricing proposal will consider magnitude of therapeutic improvement, and therefore participants considered that it would be important for NICE to ensure that the Committee considers criteria which are consistent with the Government's proposal.
9. **End of life.** Participants noted that although extension to life is already weighted through the end-of-life criteria, it is quality of life which is often more important to patients. They considered this to be particularly important when the impact of recommending a technology could mean

that the provision of palliative care is displaced. There were several participants who expressed their dissatisfaction with the way in which the current end-of-life criteria were added to the Methods Guide as a supplement. Some participants suggested that the existing QALY weighting for end-of-life treatments should be removed from the methods guide. Instead, it should be replaced by a more evidence-based approach (such as weighting based on disease severity).

10. Patient preference and the process of care. Two of the five groups thought that patient preference should be weighted as it is not reflected in the QALY gains. This could include changes to the delivery of care (to reduce anxiety to the patient, or treatment which fits in better with family life), or a weight which is attached to a certain type of treatment (for example less invasive treatments compared with standard practice). Participants thought that although these issues are captured during the Committee's deliberations, they could be weighted to ensure that they are consistently addressed in all appraisals. This would also bring the Committee's approach in line with the NHS operating framework which places a lot of emphasis on patient preference and process of care are.

How should criteria be defined?

Participants expressed confusion about how the criteria should be defined. However, they were unanimous that once criteria are selected they should be clearly defined along with their trigger points.

The groups considered how QALY weighting would be applied to severity of disease. One suggestion was that an audit of all technology appraisals conducted to date could be undertaken to rank diseases by severity. An arbitrary cut off could then be applied so that the top 5-10% of diseases were considered to meet the criteria for severity. Another suggestion was that a study should be commissioned to elicit societal preferences for weighting different severities of disease. There was also a suggestion that people with health states which are considered to be 'worse than death' were a special case which required special consideration for QALY weighting. Some

participants considered that the criteria for disease severity should be binary (yes or no) rather than a continuous scale in which different diseases could have a different weighting applied based on differing severities.

When discussing how additional QALY weighting would be applied to children, several issues were raised. For example, how should children be defined and should there be a different weighting applied to different ages of children? Some participants were concerned that a child aged 17 years might receive a different QALY weighting to that which would be applied when they turn 18 and become an adult. Participants concluded that if children were to have a different QALY weighting to adults, then it would be important to first commission a study to evaluate the societal preference of QALY weighting in children.

3.3 Determining weights for specific criteria and incorporating them into the appraisal process

Who should decide on the QALY weights and how should they be elicited?

Some participants considered that the decision over whether or not there should be QALY weighting was not for them to answer. Many suggested that it should be a political decision (by the Health Minister for example) which is then left to NICE to implement as appropriate. Some participants questioned whether the Committee was qualified to make the social judgements that may be required should QALY weighting be implemented.

Many participants acknowledged that studies to estimate the weights for specific criteria are already being undertaken (by the University of Sheffield) to support the government's Value Based Pricing proposal. It was suggested that this research could also be used to inform the QALY weights which could be applied by the Committee when assessing new technologies.

Four approaches to determining QALY weights arose from the discussions:

1. **Using population-level preferences**— this view emphasised the need to reflect the views of society based on a random sample of the general population. However, a number of weaknesses with this

approach were identified including lack of consistency in the results depending on how questions are asked, and which survey tool is used. To minimise systematic biases, one recommendation was to conduct higher quality surveys to yield more reliable responses.

There was consensus that the most appropriate method to elicit population preferences was not clear and that all the different methods currently available have limitations. However, the Discrete Choice Experiments (DCE) approach was considered to be the least worst option despite issues around consistency and bias. One view expressed by participants was to accept the imperfections of the DCE, in the same way that the EQ-5D is used to estimate quality of life despite its limitations.

Of note, participants recognised that some societal preferences may be undesirable for NICE to adopt, for example, if they were considered to promote inequities.

2. **Using a small group of expert people (not manufacturers)** – it was considered that this approach would result in more consistent judgements but that any expert elicitation would need to be justified, transparent and subject to consultation and negotiation. The question of whether a small group of people had the authority to determine QALY weights was also raised by participants.

There was no clear opinion regarding how the QALY weights would be elicited using this approach.

3. **Political decision** – this view emphasised that weights should be handed down to NICE from politicians, namely the Secretary of State, because participants considered that the Secretary of State is the only person with the political mandate to make the decision. This view expressed the need for politicians to be explicit when they prioritise one group over another.

4. **Mixed approach** – because of the limitations associated with using a single approach, a common view was to have a mixed approach consisting of:

- public preference followed by expert adjustment; or
- a political decision informed by general population preferences

How should QALY weights be incorporated into the appraisal process?

Participants noted that the feasibility of incorporating QALY weights into the appraisal process would be dependent on obtaining reliable, validated QALY weight estimates, the number and complexity of criteria for consideration; and whether the criteria are binary or continuous variables.

Citing DH-EEPRU's work on burden of illness with respect to the Value Based Pricing consultation, participants noted severity as the most likely (measurable) criteria that could be included as a QALY weight for future technology assessments. Due to the nature of a decision analytic model, a severity QALY weight may be a relatively simple addition to the model. Participants noted that for criteria such as unmet need and innovation, obtaining a reliable QALY weight may be difficult; therefore, these may be better suited as context-specific discussions through the deliberative process.

Participants cited concerns over a potential increase in the complexity of the analytical models if QALY weights are formally incorporated into them, and the potential shift towards patient-level modelling, which would impose resource issues to ERGs. For example, if the criteria are continuous variables (for example, for age) the QALY weight would, hypothetically, be applied according to the distribution of simulated patients entering the model, to estimate an unbiased estimate of the mean weighted QALY gain. This would increase the level of work and complexity of the model, and some participants voiced concerns that it may complicate interpretation and reduce transparency. Many participants cited it would be preferable if the QALY weights were binary (e.g. child versus adult), as it would be more feasible to incorporate them into the economic analysis and present to the Committee as

a secondary or sensitivity analysis (not the primary analysis). Participants considered that if the QALY weights were small (i.e. close to 1), then there would be insufficient benefit gained from incorporating formal QALY weights into the economic analysis when considering the additional burden for sponsors, NICE, review groups and the Committee.

Participants noted that particular methodological issues may arise when trying to incorporate weights into a decision analytic model (survival analysis/uncertainty); however, several participants considered that these should not be seen as a reason to avoid formally incorporating QALY weights.

Participants generally felt that QALY-weighted analyses should be presented to the Committee for deliberation, where the Committee would comment on the validity of the QALY weights, its impact on the technology, and consider the other 'non-modelled' criteria (ex. innovation, unmet need, etc...).

Participants on the whole agreed that the deliberative process was required to ensure scientific accuracy, including that the weights have been appropriately modelled, along with considering other non-modelled criteria.

Several groups noted that there is a preference for a simple process and a need for more transparency into how additional criteria would be discussed and reported by the Committee. Some participants suggested that incorporating QALY-weights would add to the complexity of the process, potentially reduce transparency and possibly contribute to inconsistency in how it would be assessed by the Committees. It was not clear if people felt strongly whether a deliberative consideration of all QALY weights would be more or less transparent than an analytical consideration of all QALY weights. One participant commented that if QALY weights are to be included, then this should be done correctly, via the analytic model (and solve any methods issues), rather than doing it simply and incorrectly. One participant cited that previous NICE appraisals set a precedent for future QALY-weightings and unspecified weights can be inferred through case law. Participants cited the example of end-of-life criteria, but noted that this criteria and weight has weak theoretical underpinnings.

Participants encouraged NICE to ensure that the QALY weight applied to end-of-life treatments is explicitly described in the NICE methods guide.

Opportunity cost issues arising from QALY weighting

....if QALY weights are formally incorporated into the process

There was general agreement among participants that if QALY weighting is formally incorporated into the cost-effectiveness framework, then it should also be reflected either formally or informally in the opportunity cost of displaced technologies.

There was also agreement among participants that the best way of dealing with the opportunity cost issues arising from QALY weighting was to apply weights symmetrically such that both QALY weights above and below 1 can be incorporated. It was agreed that, within the context of a fixed NHS budget, this was the optimal approach; that is, by allowing for higher QALY weights this will implicitly mean that more technologies will need to be displaced from the NHS budget to account for the higher 'value' being placed on the health benefits (QALY weighting > 1) for specific patient groups and so therefore, some technologies will need to be downgraded for other patient groups (QALY weighting < 1). By only applying positive QALY weights to specific technologies, the opportunity costs may exceed the benefits (QALYs) gained for a given NHS budget.

There was discussion among participants about some of the challenges of applying QALY weights symmetrically, in particular in situations where QALY weights have not already been applied to existing technologies which are to be displaced within the NHS. In addition, some participants acknowledged that the relative cost-effectiveness of many technologies, including those that should be displaced, are unknown and that some cost-effectiveness evaluations of technologies that are to be displaced (including any additional QALY weighting for these technologies) may need to be undertaken. This may involve considerable time costs. There was also general acknowledgement that NICE does not currently have a formal system in place to evaluate which technologies (currently funded in the NHS) should be displaced/disinvested when more cost-effective technologies are introduced. It

was also noted that if NICE provided more explicit advice on technologies which should be disinvested, this would be helpful for PCT-level decision-making.

Several concerns were raised about how QALY weighting will be implemented. Specifically, participants noted that negative QALY weights will need to be applied to some patient groups (e.g. healthier patients, less severe disease/illness) which will be politically unpopular. There were also concerns raised about the technical difficulties involved in ensuring that the positive and negative weights are equally offset across all technologies that are appraised (that is, that it must sum to zero).

Overall, there was little support among participants for making explicit changes to the current threshold in order to allow for formal consideration of QALY weighting into the cost effectiveness framework. There were some participants who argued that if the current cost-effectiveness threshold was adjusted downwards or upwards to reflect the current NHS budget (that is, to reflect the opportunity cost of technologies displaced by new, more cost effective technologies), this may reduce the need for QALY weighting. For example, if the threshold was adjusted downwards then it may be possible to only apply positive QALY weights where necessary. However, this would require full knowledge of the costs and QALYs from all technologies funded within the NHS, which currently does not exist. There was general awareness that the current cost-effectiveness threshold has no formal empirical basis and that there is ongoing research (Claxton et al., York University) that will attempt to derive a formal, empirical estimate of the threshold.

.... if QALY weights are reflected in a less formal manner?

Some participants argued that if QALY weights are applied less frequently (that is, only in special cases) then the issue of opportunity cost and applying simultaneous negative weights may be less important, as the overall impact may be negligible and, as a consequence, a formal adjustment to the cost-effectiveness threshold would not be necessary. Subsequent to this argument, it was noted that 'end of life' criteria, which involves QALY weighting are applied frequently in many appraisals for advanced cancers.

There was concern among some participants that, if done on a deliberative, case-by-case basis, then judgements on QALY weights made by each individual committee may lack transparency and consistency. There was some acknowledgement of a trade-off between a more formal, transparent approach (that could be applied on a uniform basis across all committees) and a more flexible, ad-hoc approach (which would allow individual committees to estimate their own QALY weights). The general consensus was in favour of the former approach.

4 Key issues for consideration by Working party

1. Can the Methods Guide describe how the QALY weights will be applied to additional criteria and what influence they should have on decision making?
2. Will it be possible to include in the Methods Guide an explicit list of criteria together with their respective weights?
3. What are the benefits of formally including QALY weighting into the appraisal process through an algorithmic approach rather than just through deliberation?
4. How should a decision be made about which criteria should have QALY weights applied to them?

5 Authors

Prepared by Fiona Rinaldi on the basis of workshop feedback from

Helen Starkie, Claire McKenna, Jeshika Singh, Eleanor Donegan, Jon Minton, Zoe Charles, Jon Tosh, Kumar Perampaladas, Manuel Gomes, Matthew Dyer, whose contributions are gratefully acknowledged.