

# Summary of a workshop to inform the research agenda on improving cross-sector comparisons using QALYs and other measures

## Introduction

From its inception in 1999 NICE's decision making has used the health-related quality of life adjusted life year (QALY) which captures both the length and aspects of quality of life (QOL) that are related to health. Historically the EQ5D questionnaire has been NICE's preferred method to calculate the QALY because it was the only one to have a UK value set (tariff) enabling it to be used in cost-utility analysis.

Whilst there are inherent limitations in all available measures, the role of the QALY and the EQ5D in national access decisions has focussed scrutiny. The challenges have included a lack of sensitivity and that not all the outcomes that are important are captured, particularly for public health and social care interventions. An intervention may be financed by one sector and have additional benefits in others which are not captured and there is no consistency in decision-making. Therefore methodological development is required.

To inform the future research agenda addressing cross-sector decision making, a meeting was held at the MRC on 14<sup>th</sup> July 2015 attended by representatives of various academic institutions, patient organisations, NICE, the Department of Health, NIHR, health charities, What Works in Wellbeing, the Cabinet office and the Medical Research Council. There were presentations on the MRC research grant process, NICE's approach to science policy development and current difficulties faced by NICE. An overview of the current methods available for cross sector benefit measurement commissioned by the MRC-NIHR was presented<sup>i</sup> and a summary given from the social care perspective. The presentations were followed by discussion.

During the meeting, there were a number of emerging themes that can be categorised into three overlapping topics:

1. **Foundational questions of valuation and decision-making** - the underlying value judgements that need to be made to construct and utilise a benefit measure suitable for decision- making.
2. **Methods of benefit measurement** - the type of benefits NICE should prioritise over others given its remit and the broad methodological approaches available to measure these benefits.
3. **The appropriate decision rule / reference case** - how the benefit measure(s) should be applied in decision-making.

Research questions under each theme were identified with recognition that the topics are inherently interlinked. For the purposes of updating the MRC-NIHR highlight notice this document presents those research questions relevant to the method of benefit measurement. The broader themes and research questions have been included in Appendix 1.

The aim of this document is to capture the viewpoints held by these stakeholders not to endorse any particular path.

## **Summary of workshop themes relating to the MRC-NIHR highlight notice**

### ***Empirical evidence of the properties of benefit measures***

1. There have been a number of research studies that have explored the properties and sensitivity of the EQ5D because it has become linked to national decision-making. Some work has also been done on other measures but not to the same extent.
2. Each of the measures has a different conceptual underpinning and captures different constructs, either directly or indirectly. The relationship between the different constructs has not been fully described. Each construct has different issues associated with it and no consensus has emerged about the correct construct to use in different circumstances.

### ***Potentially relevant initiatives for this work***

3. The 'Budapest Initiative' (UNECE/WHO/Eurostat Task Force on measuring health status) is developing internationally comparable estimates of population health status. This requires the development of a common set of instruments to measure health status in its multiple dimensions.  
<http://www.unece.org/index.php?id=28480>
4. The COMET initiative (Core Outcome Measures in Effectiveness Trials Initiative) is identifying agreed sets of outcomes for different diseases which are recommended for use in research. <http://www.comet-initiative.org/>

### ***Mapping between measures***

5. There has been work using an Australian dataset to map between different wellbeing measures and constructs.
6. There are a number of existing data sets that have been created and could be utilised.
7. Preliminary work has been commissioned from the Department of Health's Policy Research Unit in Economic Evaluation of Health and Care Interventions.

### ***One measure or many measures?***

8. There is a persistent tension (or trade-off) between:
  - a. The sensitivity of a measure to pick up subtle or gross changes in a particular state (i.e. health or wellbeing).

- b. The ability of a measure to be broadly applied across either disease states (like the QALY) or measure broader categories of social benefit (for example learning, culture and self-actualisation).
- 9. There are many disease specific measures that are very sensitive to changes in a single disease area. Use of a disease-specific measure makes comparison across diseases difficult.
- 10. Empirically, there appears to be a choice between adopting:
  - a. One measure that at least partially captures everything deemed to be of importance. This could potentially provide the basis for exchanging between measures.
  - b. Many measures and a means of exchanging between them.
- 11. There are some concerns about the ability of current psychometric methods used to map between measures and the ability of these methods to measure a concept such as wellbeing. Whilst outputs may have a numerical relationship that can be described, this may not be meaningful or valid in real terms.
- 12. As it is impossible to capture all possible benefits, there is some debate about which benefits NICE should prioritise over others. There are a lot of options for example social care, public health, education, adult education, employment outcomes and culture.

### ***Requirements of an ideal measure or measures***

- 13. The measure should be applicable to a range of groups and reflect their preferences. For example across children/ young people, older people, and ethnic groups.
- 14. The measure should be applicable across a broad range of sectors reflecting diverse benefits. For example, social care, public health, education, adult education, and culture.
- 15. The measure should be sensitive to changes in the underlying construct (for example, health or wellbeing).
- 16. The measure should be able to capture both of quality and length of life.
- 17. The measure should capture how the benefit (for example health or wellbeing) is displaced 'at the margin'. i.e. it should be possible to estimate how much of benefit is lost when the available health budget contracts.

18. The measure must provide a basis for defensible decisions for NICE; it must be amenable to a transparent decision rule.
19. The measure needs to be extrapolated over long time horizons simulating the lifetime of an individual.
20. It should be quick to fill in and simple to analyse.
21. It should not be too complex.

### ***Psychometric methods***

22. Whatever benefit instrument is developed it will rely on psychometrics (the measurement of psychological concepts). It was suggested that current methods are not sophisticated enough to do this and we need more advanced methods.

### ***Conceptual clarity***

23. It was mentioned that terms exist for both 'subjective' and 'objective' wellbeing. It was questioned whether or not these are distinct concepts and whether a differentiation is even possible.
24. It was questioned whether it is some idea of 'health-related wellbeing' that the health service should prioritise over for example 'work-related wellbeing'. It was questioned whether these are concepts that could be differentiated.
25. It was noted that the usage of the terms 'quality of life' and 'wellbeing' have changed over the years and that disagreement in usage likely persist. It was suggested that 'quality of life' is more neutral than 'wellbeing'. In the evaluation of quality of life there is more of an external reference, less of an internal reference and less fluctuation over time.
26. There is debate about what the relationship is between wellbeing and mental health. Empirical studies suggest that overlap exists between the two constructs.
27. The variation in definitions and terminology was discussed. It was noted that it would be helpful for consensus to be reached but it was unclear whose responsibility this should be.

### ***Moving forward***

28. It was suggested that there are limitations to what can be known about these benefit measures before they are actually developed. Wellbeing measures may need to be developed before they can be investigated empirically.

29. The potential impact and acceptability of moving from a system of health to wellbeing was noted. For example if someone has adapted to a poor 'objective' health state and has high wellbeing are we willing to accept that this person has a low need for interventions?<sup>ii</sup>

## **Summary of potential research topics**

- A. What are the benefits that NICE should capture in each of the different sectors<sup>1</sup>? What measures are available to capture those benefits and how do those measures they relate to each other conceptually?
- B. What is the difference in performance, acceptability and impact of the use of one comprehensive measure or a number of approved measures whose relationship can be mathematically described?
- C. Is mapping between specific measures a viable solution to benefit measurement across-sectors?
- D. How can inter-operability between measures be achieved? Can we use the existing data sets?
- E. What properties should a single measure or measures have to address the challenges of decision-making across sectors which NICE faces? Is it feasible for such a measure to be created?
- F. What are the conceptual distinctions and relationships between the different approaches and what is the impact of using them for allocation decisions? For example, health, subjective wellbeing, objective wellbeing, quality of life and mental health.
- G. Are the current psychometric methods sufficient to measure difficult to observe constructs such as wellbeing? What methods are more viable?
- H. What is the difference in performance and acceptability of a system that measures benefit measurement across-sectors and one that translates all benefits into monetary terms? For example, using contingent valuation, implied willingness to pay via displacement or monetising experience. This should include a consideration of how equity and inequalities could be taken into account.
- I. What is an appropriate 'upper anchor' for each of the different conceptual approaches?

---

<sup>1</sup> For example social care, public health, education, adult education, transport, the criminal justice system and culture

## **Appendix 1 Summary of workshop themes not related to the scope of the MRC-NIHR Highlight notice**

### ***Underlying basis of decision making***

1. In calculating the benefit created, we should continue with calculating total benefit by multiplying a measure of benefit (e.g. wellbeing or health related quality of life) by time in that state.

### ***Monetising benefits***

2. It was noted that the treasury ultimately deals in monetary units, as do other government sectors such as transport.
3. A distinction was drawn between the different methods translating benefits into monetary units:
  - a. Stated preference willingness to pay (aka contingent valuation). This involves asking people hypothetical questions about how much they are willing to pay for some outcome. Some scepticism about this method was noted.
  - b. Implied willingness to pay via displacement. This involves empirical estimation of the marginal productivity of the system which implies a relevant opportunity cost. This is the basis for the York threshold work.
  - c. Monetising experience. It was noted that this is potentially theoretically valid means of trading between diverse aspects of benefit.
4. A noted advantage of monetising benefits is that theoretically this provides a framework for increasing or decreasing budgets.

### ***Decision rules***

5. The current dual narrative in monetary valuation between a 'willingness to pay' per QALY of £60,000/QALY and an 'ability to pay / displacement' of £15,000/QALY was discussed.
6. The current cost per QALY rule:
  - a. Is considered relatively simple and this is a positive aspect of the current cost/QALY approach.
  - b. Should be kept separate from whatever measure or measures move 'beyond' the QALY'.
7. How decision makers are actually going to use the new measure(s) must be taken into account during their development.



### ***How should public values be understood? (welfarist vs extra-welfarist debate)***

8. There was disagreement about whether it is only the sum of individual benefits (wellbeing, experiences, etc.) which count or whether there are concerns over and above impacts on individual benefits. The implications of this question was noted for wellbeing:
  - a. Adaption: If an individual is satisfied with their condition (health state) do we value their health over and above its effect on their wellbeing? For example, are we happy not to help someone with poor health but high wellbeing?
  - b. Valuation: In order to compare how desirable some health/wellbeing states are compared to others should we continue to use general public values or switch to asking the people who actually have the disease? Argued that a rational taxpayer should take account of the views of the recipients- as it may be them in the future.
9. It was noted that preferences may be meaningfully different across demographics such as age, income, ethnic group etc.

### ***Complex relationships: Productivity, education, wellbeing, health etc.***

10. These complex relationships should be captured in a full cost effectiveness analysis. There were some concerns about how well these complex effects can be captured in single or multiple indices.
11. There is a need to understand the links between:
  - a. Wellbeing and health; how wellbeing affects health, how health affects wellbeing.
  - b. Wellbeing, productivity and GDP; how wellbeing affects productivity and how productivity feeds into GDP. Incorporating this into a decision analysis may require values for utility of consumption and effects on taxation rates.
12. The potential to double-count across departments was noted and it was stressed that clarity was needed about what is being valued. For example, if education affects GDP and the increase in GDP is taken into account we may not need to measure preference/wellbeing associated with the education itself.

## ***Productivity***

13. It was argued that productivity can be captured on the cost-side of cost-utility analysis. For example, instead of measuring the 'benefit' provided by being more productive, this could be captured by reducing the 'cost' of the intervention as productivity increases.
14. There were questions raised about who should value productivity and how it should be valued.

## **Emerging research questions**

### **Foundational questions of valuation and decision making**

- A. Should total benefit be calculated by multiplying a benefit measure (for example, wellbeing or health related quality of life) by time in that state? Are the benefits equal in different circumstances for example during childhood and at the end of life?
- B. In order to compare how desirable some health/wellbeing states are compared to others should the relative valuation of health/wellbeing states be based on people responding to surveys (stated preferences) or some other method such as an absolute psychological measure of health/wellbeing?
  - a. If preferences are utilised to rank health states, should they come from the stated preferences of the general public or the people affected by the decision?
  - b. If preferences are not used, what other options are available and how viable are they?
- C. Preferences change over time and vary between different groups (for example by age, gender or ethnicity). Can and should a system be developed that allows and takes account of that heterogeneity?
- D. Is it only the sum of individual benefits (wellbeing, experiences, etc.) that count? If an individual has adapted to their current condition (health state) and is satisfied do we value their health over and above its effect on their wellbeing?
- E. How should inequalities and equity feed into decision making?

## The appropriate decision rule/reference case

Compared to health, the appropriate decision rules in social care and public health are less well defined and agreed upon. This section seeks to address the need for useful decision rules within and across sectors.

- F. What form should the decision rule(s) take for each method of measuring benefit; mapping between a suite of measures, one comprehensive measure, and monetising benefits?
- G. What should be the relevant components of the NICE reference case i.e. how should the cost-benefit calculation be structured?
- H. What should be the underling theory and justification for the decision rule or rules?
- I. What are the requirements of an ideal decision rule? Additionally, what are the sufficient properties that a decision rule must have to address the problems of decision -making across sectors which NICE faces?
- J. What must be taken into account when comparing possible uses of different combinations of benefit measures and decision rules?

---

<sup>i</sup>Brazier, Tsuchiya. Improving cross-sector comparisons using QALYs and other measures: a review of alternative approaches and future research. <http://www.mrc.ac.uk/documents/pdf/improving-cross-sector-comparisons-using-qalys-and-other-measures-a-review-of-alternative-approaches-and-future-research/> [Accessed 11 August 2015]

<sup>ii</sup> In practice these approaches would be unlikely to be used on an individual level and are more likely to be used in incremental decision making across populations. The question then becomes: if a certain population has been shown to adapt better to a poor “objective” health state than another population with a better “objective” health state then are we willing to accept that the former population has a lower need for intervention (given effective interventions)?