

BRIEFING PAPER

DEPARTMENT OF HEALTH PROPOSALS FOR INCLUDING BURDEN OF ILLNESS INTO VALUE BASED PRICING: A DESCRIPTION AND CRITIQUE

NICE DECISION SUPPORT UNIT

July 2013

Alec Miners,¹ John Cairns¹ and Allan Wailoo²

¹ London School of Hygiene and Tropical Medicine (LSHTM)

² School for Health and Related Research (ScHARR), University of Sheffield

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

Tel (+44) (0)114 222 0734

E-mail dsuadmin@sheffield.ac.uk

www.nicedsu.org.uk

ABOUT THE DECISION SUPPORT UNIT

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

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

The production of this document was funded by NICE. The views, and any errors or omissions, expressed in this document are of the author only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

CONTENTS

1. BACKGROUND	4
2. THE DEPARTMENT OF HEALTH PROPOSALS AND BROADER FRAMEWORK	5
3. THE REFERENCE DATA SET	7
4. DECISION SUPPORT UNIT REMIT	9
5. TERMS OF REFERENCE TO NICE FROM MINISTERS	9
6. THE DEPARTMENT OF HEALTH'S PROPOSALS FOR BURDEN OF ILLNESS	10
6.1 The proposed definition of BoI	10
6.2 What is the DoH proposing for BoI?	11
6.3 Measuring BoI	12
6.4 Valuing the contribution of BoI to a specific treatment	12
6.5 Valuing displaced treatments	13
7. DSU EVALUATION OF THE BURDEN OF ILLNESS PROPOSALS	16
7.1 Definition of BoI	16
7.2 BoI metric and measurement	17
7.3 BoI valuation	19
7.4 End of Life	21
7.5 Valuing displaced treatments	22
8. POSSIBLE CONFLICTS WITH OTHER ELEMENTS OF NICE'S TECHNOLOGY APPRAISALS PROGRAMME METHODS GUIDE	22
9. THE WAY FORWARD?	24
10. FURTHER RESEARCH	25
11. REFERENCES	26
APPENDIX 1: CLARIFICATION ON ASPECTS OF THE THRESHOLD PROJECT PROPOSED TO BE USED WITHIN VBP	27

FIGURES

Figure 1: Burden of Illness according to Length- and Quality-of Life gains	7
Figure 2: Profile Information	11

1. BACKGROUND

All health care interventions referred to NICE's Technology Appraisal Programme are subject to an economic evaluation, in order to help the Appraisal Committees determine whether they are cost-effective and should be funded by the NHS. There are a number of methodological considerations that must be considered within any economic evaluation. For example, the most appropriate source of estimating relative treatment effects and the time horizon over which a technology should be evaluated.

One of the difficult tasks that confronts NICE is to be consistent across technology appraisals in terms of decision-making. One of the approaches it has developed to help with this is a 'reference case', in which the Institute specifies a set of methods that it believes is the most appropriate given its objectives.

One of the defining features of NICE's current reference case is that each additional quality-adjusted life-year (QALY¹) should receive the same weight regardless of any other characteristics of the people receiving the health benefit [1]. However, the Appraisal Committees have the discretion to consider a different equity position, and may do so in certain circumstances when instructed by NICE's Board. The clearest existing example of this is to 'life-extending treatments at the end of life' (more commonly referred to as the 'end of life' (EoL) criteria), which is applicable when the following three criteria have all been met:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment
- The technology is licensed or otherwise indicated, for small patient populations normally not exceeding a cumulative total of 7,000 for all licensed indications in England

The EoL criteria aside, the current process of assessment of new treatments can be broadly thought of as determining whether the benefit to patients of a new technology, measured in QALY gains, are sufficient to offset the losses to patients elsewhere in the NHS when funds are re-allocated to the treatment under evaluation. If they are, then a treatment can be considered cost-effective (as it generates positive net health benefits), if not, then the introduction or continued use of an intervention at hand is much less likely to be recommended for use on economic grounds (as the net health benefits are negative). In principle, the decision regarding cost-effectiveness can be made by

¹ QALYs are typically calculated by combining information on health-related quality-of-life with evidence on survival. One year of perfect health is equivalent to 1 QALY, two years in half perfect health are also equivalent to 1 QALY.

comparing an estimated incremental cost-effectiveness ratio (ICER) for a given intervention, against a threshold value of a QALY, as the latter represents the value of the notional displaced QALY in the NHS, or opportunity cost of a decision and the amount of health that is offset. In other words, the cost-effectiveness threshold defines the tipping point at which positive / negative net health benefits are defined. In theory, this approach ensures that new treatments do not displace more health gain than they provide, and will lead to decisions which do not diminish the value of benefits gained from the NHS budget, subject to two conditions:

- That all QALYs provided by treatments, or displaced elsewhere in the NHS, are of equal value to society (ie a QALY is a QALY is a QALY)
- That only patients are affected, through the health gains (or losses) from new (or displaced) treatments – and that society is not concerned with any other wider impacts

Broadly speaking cost-effectiveness is taken into account by Appraisal Committees by assessing the robustness of submitted estimates, comparing them to the Institute’s stated cost-effectiveness threshold range and by taking into account other issues that are considered important but are not necessarily captured in the cost-effectiveness assessments. The latter can be thought of as more of a ‘deliberative process’ given remaining issues of importance.

2. THE DEPARTMENT OF HEALTH PROPOSALS AND BROADER FRAMEWORK

This Working Party has been assembled to debate the proposal from the Department of Health (DoH) that these two conditions do not always hold. More specifically, the DoH is suggesting that in some circumstances, the evidence suggests that society places greater value on QALYs gained than in others, as part of the wider process of value based pricing (VBP). These two circumstances are:

- **‘Burden of illness’** (BoI) – society places higher values on QALYs gained by individuals with relatively high ‘burden of illness’.
- **‘Wider societal benefits’** (WSB) – treatments not only impact on an individual’s health, but also on wider considerations such as their ability to return to work / contribute to society. These wider aspects are not currently captured in NICEs assessments of cost-effectiveness.

The main purpose of this document is to expand on the proposal for BoI. However, we believe it is important to understand the entire framework when deliberating the BoI element of VBP, hence we also include some details as they currently relate to WSBs.

The WSB framework proposed by the DoH estimates the net WSB as a function of age, sex, International Classification of Diseases (ICD) code and quality of life (QoL). This is the difference

between the value of production by an individual with those characteristics and the value of their own personal consumption. WSBs are a financial value that is not considered to be commensurate with costs incurred within the health system. A value of £60,000 per QALY is the exchange rate between non-health costs (WSBs) assumed by the DoH. This then provides a QALY equivalent that can be added / subtracted from the overall QALY gains from treatment and their NHS costs in order to estimate an adjusted ICER:

$$\text{adjusted Q} = Q + \left(\frac{\text{WSB}}{\lambda_2} \right)$$

Where Q is QALYs, WSB is the monetary value of Wider Societal Benefits, and λ_2 is the exchange rate between non-health costs and QALYs. The value in £s per month is calculated both for the comparator treatment and the new technology. Ideally this requires an estimation approach that reflects both the non-linear relationship between age and WSB and the fact that age and QoL will change over the time where QALYs accrue. This implies that a patient-level model of the technology itself would be required. Anything less than this is an approximation and the degree of ‘accuracy’ of the estimate is likely to be unacceptably large (in our opinion). The WSB-adjusted QALYs for the mean of a cohort of patients is not equivalent to the mean WSB-adjusted QALYs for a cohort of patients. However, NICE rarely, if ever receives cost-effectiveness analyses that are true patient-level simulations in the sense that they reflect the full heterogeneity of the patient population.

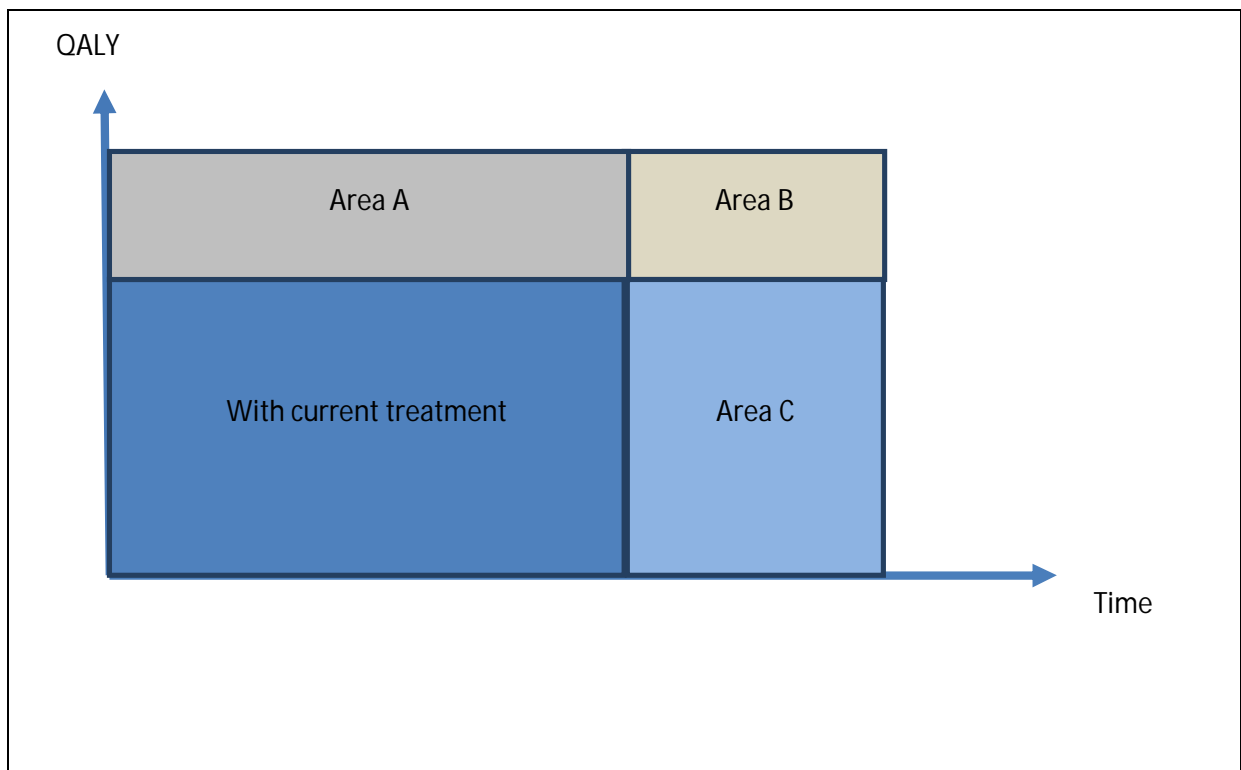
The DSU has previously explored the extent to which results vary according to various simplifications and approximations that could be made either using the summary information that results from a cost-effectiveness analysis (CEA), i.e. mean costs and QALYs for each arm, or by incorporating WSB estimates within the cohort models that are typically submitted. Whilst this work was by no means exhaustive, our tentative conclusions were that there would be a substantial risk of bias from any of the approaches we tested.

3. THE REFERENCE DATASET

In order to avoid the need to make such approximations within the submitted cost-effectiveness models, an alternative approach has been suggested by the DoH which utilises an external data source rather than the submitted dataset. This “Reference Dataset” is based on the UK element of the WHO Global Burden of Disease project [2]. Indeed, the data source is used to estimate a number of parameters in both the BoI and WSB components of the VBP proposal. The Dataset was used in the recent ‘Methods for the Estimation of the NICE Cost Effectiveness Threshold’ study [3].

The data is divided into 1,281 disease areas (ICD codes) and further divided into broad age (8 categories) and gender categories. This shows the distribution of patients by age and sex, for each individual disease area. This is the level of aggregation at which the data are supplied by the WHO: this is not an analysis choice. The data within each unique ICD, age and sex category provides a) the average QoL b) the overall burden of illness measured in QALYs both as an aggregate figure and broken down into i) loss due to lower QoL and ii) loss due to length of life (LoL) and iii) the typical LoL gain from treatments in this disease category. Figure 1 below illustrates this (see Appendix 1 for a more detailed explanation of the Dataset).

Figure 1: Burden of Illness according to Length- and Quality-of Life gains



In the Figure above, the difference between QoL and LoL with and without the condition in question is illustrated. It can be seen that this burden of illness comprises i) Area A, pure loss of QoL ii) Area

C, which comprises pure loss of LoL and iii) Area B which is loss due to both. It is our understanding that area B has been categorised as loss due to LoL i.e. combined with area C, and that this is the case both for defining the burden of illness associated with each disease area AND the proportion of QALY gains due to life extension from therapies in this disease area.

This Reference Dataset is used in four places in the overall VBP framework as proposed by the DH:

- i) Within the disease area for technology being appraised: To estimate the burden of illness, how that is divided according to LoL and QoL, AND the proportion of gains from therapies which are generated from LoL improvement. This last step is important – the clear alternative to using the reference dataset here is to extract this information from the submitted cost-effectiveness model.
 - ii) Across all disease areas: to estimate the mean burden of illness, how that is divided up according to LoL and EoL, and how the gains from treatments in each of those disease areas are divided up between EoL and LoL gains.
 - iii) Within the disease area for the technology being appraised: To estimate the distribution of patients by age, sex and QoL which in turn calculates the WSB gain from treatment. Avoids the need to do it in the cost-effectiveness model (which is usually based on mean age). This assumes many things:
 - a. The patient population affected by the technology is the same as the entire ICD coded population. This is unlikely to be the case.
 - b. The absolute treatment benefit is exactly the same for all patients irrespective of their age or sex within the distribution.
 - iv) Across all disease areas: to estimate the mean WSB for all ICDs, given their age, sex and QoL distribution for estimation of the notional displaced QALY.
- For BOI

Bringing together all this information permits a calculation of an adjusted threshold value, against which the standard cost per QALY estimate can be judged and around which the maximum amount the NHS should be willing to pay per unit of the drug can be established.

4. DECISION SUPPORT UNIT REMIT

In broad terms, the Decision Support Unit (DSU) has been asked to review these two proposals, given the dossier of evidence submitted by the DoH (one relating to BoI and the other to WSB), and to provide the Working Party with a commentary on them (in light of the Terms of Reference (ToR, see Section 5) for implementing VBP to NICE from Ministers). More specifically, and in terms of BoI,² the DSU have been tasked to:

1. Review the proposal and the dossier from the DoH
2. Summarise key aspects of the research/evidence submitted (methods and findings) and the proposed approach for weighting for BoI
3. Evaluate the proposed approach / methods
4. Identify any issues associated with the proposed approach
5. Identify any gaps in the evidence
6. Consider impact on relevant concepts included in the Guide to Methods of Technology Appraisal 2013 (including ‘equalities’) [1]
7. Consider opportunities for use of evidence currently submitted to the Technology Appraisal Programme, or otherwise the consequences for the submission of evidence

5. TERMS OF REFERENCE TO NICE FROM MINISTERS

The methods for value assessment of branded medicines under VBP should:

1. Be applied to medicines within the scope of the VBP system, and incorporated into the methods for other categories of guidance at NICE’s discretion
2. Adopt the same benefit perspective for all technologies falling within the scope of VBP, and for displaced³ treatments
3. Be as transparent and predictable as possible
4. Be informed by the best available evidence
5. Include a simple system of weighting for burden of illness that appropriately reflects the differential value of treatments for the most serious conditions
6. Encompass the differential valuation of ‘End of Life’ treatments in the current approach within the system of Burden of Illness weights
7. Include a proportionate system for taking account of Wider Societal Benefits
8. Not include a further weighting for Therapeutic Innovation and Improvement
9. Produce guidance for patients and the NHS which describes the clinical and cost-effectiveness of the technology and its position in clinical practice

² As previously indicated, a separate briefing document is being prepared for WSBs.

³ That is, the value of a new treatment is considered net of the value of what is displaced.

While all the ToRs are key to discussing the implementation of the VBP proposals by the Working Party, and NICE, those considered to be particularly pertinent to discussing BoI have been underlined (ie criteria 2-6 inclusive).

6. THE DEPARTMENT OF HEALTH'S PROPOSALS FOR BURDEN OF ILLNESS

The DoH has put forward a dossier of evidence to NICE and the DSU in which it outlines its proposal for incorporating BoI. Its starting point is effectively the ToR from Ministers, meaning that the inclusion of BoI is a given; it is more a question of how to achieve it.

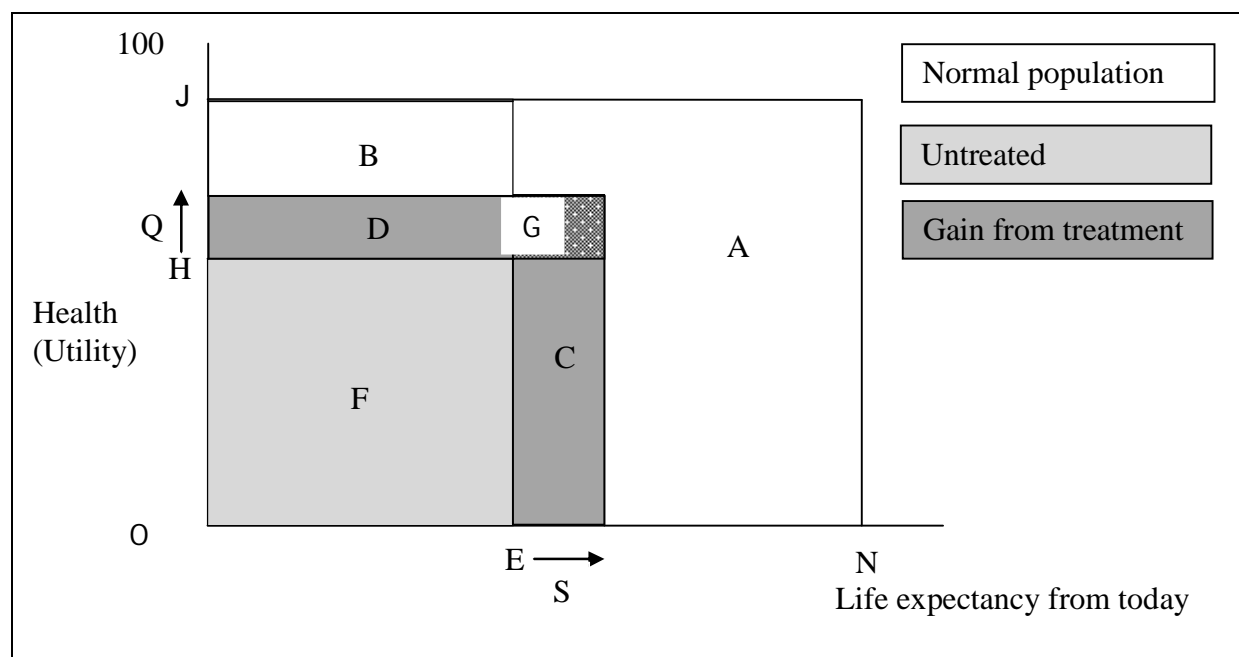
6.1 The proposed definition of Burden of Illness

The DoH's proposal is that BoI should be defined:

'.....as the number of QALYs lost by a patient because of their condition'.

More formally this is described as a person's number of projected QALYs given current treatments for their particular condition, subtracted from the expected number of QALYs given a sex-adjusted population life expectancy without the condition. It is important to note that this definition is based around assessments from the current time point, which in the context of technology appraisal is the point in a disease process at which the technology in question becomes a relevant treatment alternative. There is no single taxonomy of equity descriptions, but this approach is based on 'prospective health' (health looking 'forward' rather than 'back') since no value is placed on previous health experiences. This is illustrated in Figure 2. The origin (O) represents the point at which a decision is made (ie the point on a patient's pathway that directly relates to the issued guidance). The area F (OH*OE) represents the number of QALYs with a condition in the absence of the technology under evaluation, whereas the sum of all the areas OJ*ON = I represents full health in the absence of the condition (ie a population level of health). Therefore BoI under this definition = I-F (or equivalently [J*N]-F). It is the absolute number of QALYs that would be lost, on average, for patients for whom the new technology is being appraised given the comparator treatment within the technology appraisal, compared to patients living at full health for an average sex-adjusted life expectancy. Note there is a discussion to be had as to whether QoL in the absence of disease (J) is considered to be perfect health (ie. one) or is age-adjusted.

Figure 2: Profile Information



This Figure is an amended version of Figure 1 in Brazier et al EEPRU further analysis submitted to the DSU June 2013 [4].
The sum of all areas 'full health in the absence of treatment' = I, or equivalently $J*N$

6.2 What is the Department of Health proposing for Burden of Illness?

While the DoH's proposal is clearly supportive of the principle of BoI, they do not specify either a procedural means of achieving this (eg. by including formal evidence based weights derived from empirical studies or by a more deliberative process), or the level of weighting (premium) that should be applied. Indeed, the DoH states that they do not believe the evidence currently 'dictates' a particular weighting system and that even if the principle and definition of including BoI are clear, there is no obvious means of estimation.

Instead of a specific way forward, the DoH outline a simple, possible, approach which could, in principle, be used and is consistent with the ToR. It is important to note that the suggestion is not to directly include BoI 'adjustments' in submitted economic evaluations. Rather the adjustment is seen as an activity that can / could be undertaken following the incremental cost-effectiveness calculation. As will be further explained, suggestions for calculating these adjustments draw heavily from the Reference Dataset source.

The proposal broadly consists of three stages:

1. Measuring BoI given the medical condition at hand
2. Valuing the BoI estimate
3. Adjusting the threshold ICER to account for the value of displaced treatments

6.3 Measuring Burden of Illness

Imagine an individual is aged 20 years today (O) (see Figure 2) with a current QoL of 80 % (H) of normal. With current treatment the individual will die aged 30 (E) with no improvement in QoL. Without the condition they would have been predicted to die aged 80 (N), with this difference at 90% of (perfect) health (J). The proposal is that BoI could be calculated as $80-30 \text{ years} \times 90\% = 45 \text{ QALYs}$. Note that this is equivalent to $I (J*N) - F (H*E)$. Similarly, if a patient's life expectancy did not change, but a treatment increased their QoL by 30% for 5 years, then the BoI would be $5 \text{ years} \times 30\% = 1.5 \text{ QALYs}$.

The DoH puts forward three possible options for measuring BoI QALY losses within any technology appraisal; using the source Reference Dataset, requesting the manufacturers / stakeholders to provide this information, or a combination of the two.

6.4 Valuing the contribution of Burden of Illness to a specific treatment

Once BoI has been measured, the DoH proposes a 'simple' approach to valuing it, with each QALY loss valued at a fixed rate. For example, a 1.5 QALY BoI valued at 5% per QALY would be equivalent to a weight per QALY gained of 1.075 $[(1.5 * 0.05) + 1]$. So for example, a treatment that, in the absence of BoI weighting, produced 2 additional QALYs, would now be equivalent to 2.15 QALYs $(2 \text{ QALYs} * 1.075)$.

Taking the results from the EEPRU QALY-weighting study [4] at face level (which will be described later), they indicate that society values the two components of BoI (ie. QoL and LoL) differently; LoL is valued higher than QoL. This possibility is included in the DoH's proposal and spreadsheet by separating the LoL and QoL gains, and by applying a two- rather than one-weighting system. The proposal states that if this two-weighting system is preferred, it follows that QALY gains from treatment should also be separated out in terms of gains that are attributable to LoL (C + G) and QoL (D) (see Figure 2) and the relevant LoL and QoL BoI weights applied respectively. If it is accepted that the societal weighting for BoI differs according to whether it is attributable to LoL or QoL, then it is inconsistent to apply a single weight to QALY gains ignoring whether they are life-extending or QoL-improving. Burden of illness less treatment benefit equals remaining burden of illness.

Irrespective of the 'one-' or 'two-' weighting approach, the DoH suggests that the weights could be tapered or capped, if it is considered desirable to limit the total weighting applied to BoI. However, there are no proposals as to if, or how, these ideas should or could be implemented.

6.5 Valuing displaced treatments

Economics as a discipline is concerned with the difference in outcomes between resource allocation decisions. In the context of VBP, this is recognised in the ToR by stating that decisions on funding new treatments should (simultaneously) reflect the value of displaced treatments. In other words, it is important to recognise that when a new treatment is funded, other treatments elsewhere in the NHS are displaced and that these treatments are also associated with BoI's (and potentially WSBs).

The DoH proposal for incorporating the value of displaced treatments into the BoI-adjusted QALY estimates is to estimate the value of a notional displaced QALY using the Reference Dataset, and to use this result to 'adjust' the corresponding threshold ICER. That is, there will no longer be a single threshold per se (£20,000 to £30,000 per additional QALY) which is universally applied to all technologies. Rather, treatments that, relatively speaking, displace fewer BoI-adjusted QALY gains than they produce, should be compared against a higher threshold all else being equal (as they produce more health gains than is recognised under the current QALY calculation approach). Moreover, that technologies that displace proportionally more BoI-adjusted QALY gains than they produce, should be compared against a lower threshold, as the overall net treatment benefits are lower. Note that this means that some technologies will therefore be compared against thresholds higher or lower than the existing £20,000 to £30,000 per additional QALY thresholds depending on the balance of net health benefits they displace.

The DSU's belief is that there is logic to this approach since the threshold value of a QALY is designed to reflect the opportunity cost of a decision, or equivalently, the value of the notional displaced QALY. We illustrate this using the following example to explain this point: Imagine a cost-effectiveness threshold of £20,000 per additional QALY and an incremental cost-effectiveness ratio (ICER) for a given treatment of £10,000 per additional QALY – calculated as an incremental cost of £20,000 and an increase in QALYs of 2 (£20,000 / 2 QALYs). The key question is whether the expected gain of two QALYs is greater than the health outcomes forgone elsewhere as other NHS treatments are displaced by the additional cost [5]. This is represented by the £20,000 per additional QALY cost-effectiveness threshold. Or put another way, the threshold indicates that every £20,000 found from existing resources displaces one QALY elsewhere in the NHS. At an additional cost of £10,000 per QALY, the intervention is generating 2 QALYs whereas only 1 QALY is lost (as £20,000 is being spent), hence the treatment can be considered cost-effective as there is a net gain of 1 QALY. If the incremental cost were £40,000 instead of £20,000, then the ICER would increase to £20,000 per additional QALY (£40,000 / 2 QALYs). At this increased cost, 2 QALYs are still gained but 2 QALYs are being displaced elsewhere; the net gain to society is therefore zero. At incremental costs above £40,000, the net loss in QALYs is always greater than the net gain, meaning that the treatment cannot be considered cost-effective. Thus, if the value of a notional displaced QALY is

thought to vary by condition (because some treatments are thought to displace more BoI-adjusted QALYs than they produce or vice versa), then in principle it could be reflected in decision-making by varying the cost-effectiveness threshold.

The DoH proposes to calculate displaced activity using the Reference Dataset. This evidence combined with information on the distribution of displaced activity in theory allows the BoI associated with the notional QALY displaced at the margin in the NHS to be calculated for each specific ICD 10 code, after adjusting for age and sex distributions. It is important to note that the value of the notional displaced QALY is based on an average of all displaced NHS activity given the introduction of a new medicine, and that it does not change per individual technology – in this sense it is fixed. However, what will vary by appraisal is the number of BoI QALYs (I-F), and hence the net difference between what is generated and displaced.

Summary of key elements of the proposal

- The ToR state that BoI must be taken into account in cost-effectiveness assessments using a simple approach that reflects the differential value of treatments for the most serious conditions.
- It also states that BoI should encompass current EoL considerations.
- The value of displaced treatments should be simultaneously considered.
- BoI has been defined by the DoH as the number of QALYs lost given a current health condition. More specifically, it is the difference between an expectation of an individual's health profile given the absence of the treatment under evaluation and a population average (I-F).
- This definition emphasises prospective health, but not past health.
- QALYs lost (F) could be estimated using a single Reference Dataset, by the manufacturer or a combination of the two approaches.
- Each QALY lost could be valued at a constant rate, or as a two-part weighting system if it is believed that QoL and LoL should be valued differently.
- The weighting system could be capped or tapered to lower the impact of BoI on decision-making if thought appropriate.
- No weighting system has been put forward by the DoH.
- The value of displaced treatments should be reflected by 'adjusting' the existing cost-effectiveness.
- Technologies that displace fewer BoI-adjusted QALYs than they produce will, all else equal, face a higher cost-effectiveness threshold (that is, they will be considered to be relatively more cost-effective). Conversely, those that displace more BoI QALYs than they produce, will be compared with a lower cost-effectiveness threshold.
- Adjusting the cost-effectiveness threshold is a non-trivial task that requires information regarding likely displaced treatments within the NHS.
- Therefore, the DoH has suggested using a Reference Dataset that combines information on likely displaced treatments by ICD code, adjusted for age and sex.
- The value of a notional displaced QALY would not vary by technology appraisal, it is an average. However, the net difference between the displaced treatments by an individual medicine and the value of a notional QALY will vary; hence each technology will be compared against a unique threshold.

7. DSU EVALUATION OF THE BURDEN OF ILLNESS PROPOSALS

As previously stated, the DSUs broad remit is to review the proposal put forward by the DoH.

7.1 Definition of Burden of Illness

A key decision for the incorporation of BoI into cost-effectiveness analysis is its definition since different formulations will lead to different weights, and ultimately different priority rankings for treatments. The DoH's proposal is for BoI to be measured in terms of QALYs (LoL and QoL) and calculated as prospective losses (I-F). However, there are a number of other potential definitions of BoI, such as the 'fair innings argument' and 'proportionate shortfall' [6, 7].

The fair innings argument is based on the assumption that everyone is entitled to some 'normal' level of health achievement, either defined in terms of life-expectancy or quality-adjusted life-expectancy. Anyone failing to achieve this has lost out in terms of lifetime health, while anyone getting more than this is living on 'borrowed time'. Thus, the equity weights depend on the expected lifetime health total, which includes both past health losses (note that past health is not shown in Figure 2) and prospective health.

Proportional shortfall contains elements of the fair innings and prospective health arguments as it takes into account expected QALYs without treatment relative to expected QALYs in the absence of a condition. In this sense, it looks at past and future health and is measured as the ratio of BoI looking forward (F) divided by (imagined) perfect health in the absence of disease (I). The higher the ratio, the higher the weighting. Van de Wetering [7] gives the following example, a 30-year-old losing 1 of 40 remaining QALYs would receive low treatment priority ($1/40 = 0.025$), while a 70-year-old losing 1 of 5 remaining QALYs would receive higher priority ($1/5 = 0.2$). Thus one of the advantages of this approach is that it that it attempts to equate relative future health gains.

We note that there is at least some empirical support for all the equity definitions but only limited evidence in terms of direct comparisons of society's preferences for the various definitions. For example, Stolk [8], in a small, study found some evidence to support the 'fair innings' approach. Moreover, the more recent UK-based EEPRU QALY-weighting study [4], which was commissioned by the DoH to directly investigate this issue, appears to give some (albeit) tentative support for the proportionate shortfall approach.

The aim of the EEPRU QALY-weighting study was to estimate QALY weights for three treatment characteristics:

- (1) BoI from a medical condition given current health care interventions (i.e. QALY loss per patient from a condition due to both premature mortality measured against normal life expectancy and quality of life)
- (2) Therapeutic improvement (note we do not comment on this further as it is now excluded from the ToR)
- (3) End of life (EOL) (defined as expected survival of less than 2 years and as survival as a continuous variable).

The study was based on a discrete choice experiment (DCE) in which respondents are asked to choose between competing options defined in terms of a number of attributes and levels. The attributes in this instance were life expectancy without treatment (E), survival gain from treatment (S), health before treatment (H), health gain from treatment (Q). The results from the DCE were used to calculate the required QALY weights. A total of 3,669 respondents completed the survey online. More details regarding its results are found in the ‘BoI valuation’ section below.

As a final comment, we note that in practice, implementing the fair innings approach might be complex since while current age would be known, past health would not. That is, cost-effectiveness models start at a relevant decision point given the appraisal topic at hand. While detail is known and programmed in regarding the starting cohort’s age, QoL and sex distribution, nothing is known or is explicit about how individuals arrived at this point – this is particularly important in terms of QoL. However, since the proportionate shortfall is defined as F/I, no further information is required.

7.2 Burden of Illness metric and measurement

Irrespective of the definition of BoI, it would appear logical and consistent with NICE’s existing Reference Case for economic evaluations [1] to measure it in terms of QALYs.

The BoI QALY loss calculation as it stands requires an estimate of two parameters – health in the absence of a condition (I) and – health with the condition but without the treatment under evaluation (E). The difference between the two is the QALY loss (I-E). We note that there appears to have been general agreement at a previous NICE workshop that a population average level of health was appropriate for estimating health in the absence of a condition; this seems reasonable in most circumstances. However, this is not without problems, as is later discussed.

The DoH suggest either using the Reference Dataset or by requiring the manufactures to directly estimate E. Both approaches have advantages and disadvantages. The advantage of using the Reference Dataset is that it is available and would promote consistency of BoI estimation across appraisals. Its downside is the quality and uncertainty around the data given its complex construction

and also the extent to which it is sufficiently refined for NICE's purposes. For example, ICD 10 code D66 denotes haemophilia A. Haemophilia A is characterised by clotting factor VIII deficiency but the extent to which it impacts on an individual's life (and is therefore a burden) depends almost entirely on the level of deficiency – information that is not captured in the code. A potential consequence of this is that the BoI for severe and mild haemophilia would be under- and over-evaluated respectively. Further problems when considering stage of disease or previous treatment history can also be envisaged. NICE's scope documents are typically very specific about the position of a treatment on a clinical pathway, for example, treatments for rheumatoid arthritis following failure of two treatments, but the ICD codes are unlikely to be this refined.

An alternative could be to use the Reference Dataset as a default setting, but for stakeholders / manufacturers to submit alternative estimates if they believe it does not accurately reflect the specific circumstances. Moreover, we believe this should be a relatively straightforward task since F should be equivalent to the QALYs associated with comparator treatment arms in the economic models. This approach would also help to ensure F used to calculate the initial ICER and F used to estimate BoI, are consistent.

We previously stated that a population average level of health was appropriate for estimating health in the absence of a condition in most circumstances. However, consider the use of antiviral therapy for hepatitis C (HCV) for an individual who is an active intravenous drug user (IDU). In the absence of HCV they are unlikely to have an average population quality-adjusted life-expectancy. The same principle applies to anyone with more than two co-morbidities that are in some senses unrelated, although it is unclear how often this issue would occur. This illustration also highlights further issue with valuing future QALY gains given an individual's current health status. The current proposal implicitly assumes an individual's 'BoI status', or the factors contributing to BoI considerations, remain constant over time. That is, should the BoI estimate take into account that a proportion of currently active IDUs might cease to inject drugs in the future (ie their BoI assessment (F) might change)?

There are at least two other potential problems with the use of a population average level of health (N) as proposed. First, if the condition under scrutiny is a large determinant of a population average, then the population value will be a biased estimate when the particular cause is removed. Second, there are many socioeconomic demographics other than sex that predict future health, such as social class and smoking status. There is no clear rationale for focussing purely on sex other than perhaps for pragmatic reasons.

7.3 Burden of Illness valuation

The DoH does not propose a particular system of weights, but their evidence dossier contains a number of studies. The most substantive in a quantitative sense is the EEPRU QALY-weighting study. The most recent results (seen by the DSU in June 2013) broadly suggest that individuals value QALY gains from treatment (which is not included in the ToR), albeit at a decreasing rate. Survival before treatment is highly valued, which the authors take to indicate support for taking pre-treatment survival into account though not the specific 2-year cut-off used in NICE's current position on EoL. The results also suggest that QALY losses, BoI, as a result of QoI and LoL are not valued equally (LoL is valued more highly). However, the relationship is complex. For example, the LoL and QoL weightings were shown to be unstable with respect to expected life expectancy in the absence of disease (N). It is the DSU's understanding that this observation, and adjustments made to the analysis to account for it, have led the EEPRU authors to tentatively prefer the proportionate shortfall definition of equity (i.e. as a proportion of normal life expectancy), as defining BoI (and indeed pre-treatment survival) in these terms led to some improvements in the fit of the various statistical models. The ultimate aim of the EEPRU study was to derive an appropriate set of QALY weights. While a system of weights is presented in the report, they have not been put forward as such by the DoH. The exact reason for this is unclear, but the report authors themselves emphasise the amount of uncertainty around the results (for example, in the structure of the equations used to transform the results into weights) and the complex nature of the results relative to the ToR requirement for a 'simple' valuation approach. The face validity of the weights also requires some debate, with values of up to 4 and 5 for conditions / treatments with small QALY gains relative to large BoI assessments. We are also concerned about the possibility and appropriateness of taking and applying weights for some of the DCE attributes (BoI however defined) but not simultaneously applying results for other attributes (such as QALYs gained) since the weights are relative to each other and in theory might change given a different attribute set.

Our current understanding in terms of the status of the report is that further analyses have recently been undertaken and a new set of weights derived, that address a number of outstanding methodological concerns (personal communication with John Brazier, Professor of Health Economics, School for Health and Related Research, University of Sheffield). However, at the time of writing this report, we have not seen this analysis.

A recent study by Linley et al. [9] also set out to explore societal preferences for factors such as the Cancer Drugs Fund, NICE's current EoL criteria, unmet needs and therapeutic innovation. The study recruited over 4,000 UK participants. In terms of BoI, the results suggest support for severity of disease and unmet need (defined in terms of the availability of alternative treatment options). Similarly to the EEPRU study, preferences were also shown towards populations that gained large

improvements in health. However, unlike it, no support for EoL / survival before treatment was identified. Irrespective of these results, it is important to note that the study was not designed in a way that would allow QALY weights to be calculated. Thus, while it adds some support to the desirability of incorporating BoI, it is only useful in a qualitative sense.

The DoH's dossier also included two unpublished studies by Nord et al. [10, 11] that examine the value society places on severity, defined as a proportionate shortfall. Both studies are based on analysis of data abstracted from a review of 21 studies from 9 different countries. Their results indicate that severity is valued, although the variation in results across the individual studies was enormous. More specifically, Nord describes a severity gradient (k). See the formula below where Sdiff is the difference in severity (here defined for simplicity in terms of utility) between two health states and SV_{AB} is the relative value of health state A to B (ie the weight).

$$SV_{AB} = dU_A / dU_B \times (k \text{ Sdiff} + 1)$$

Assuming that severity can be measured, the key to 'solving' this equation, and thus to calculating the weights, is to quantify k. Nord estimates this parameter using the review and linear regression analysis on the abstracted data. The results show that when severity is 0.1, the average QALY weight was 5 to 6 compared with a value of 0.9. However there was a very large distribution in terms of individual study results and Nord comments that the values might be country specific – for the same severity span (0.1 to 0.9) the values ranged from 0.4 in the UK to 28.4 in Spain, even though the elicitation procedure was the same. Interestingly, Nord ultimately concludes that choosing a universally applicable 'best estimate' of the strength of concerns for severity is impossible, and that any 'best estimate' will need to be country specific. He also concludes, that a single statistic (here k) cannot alone describe the structure of concern for severity, which in many way highlights the difficulty of the ToR requirement to implement a simple approach to including BoI.

Nord has undertaken a similar review based study in which he generates weights to 'adjust' existing EQ-5D utilities for severity concerns. Several non-linear functions were used to estimate 'n' or the weighting that should be applied. Again, the results were very broad (estimates for n ranged between 0.4 and 32 across the individual studies) although Nord suggests that a reasonable estimate at least to start conversations should be somewhere near 2. Note that the precise impact of n on the EQ-5D values is partly dependent on the choice of utility function, which is a further source of uncertainty.

As something of a summary on this section, we again note that there are further EEPRU QALY weights to be presented. However, there is clearly a large amount of uncertainty regarding all of the empirical results such as they are. Thus, even if there is a desire to base a weighting system on the

evidence, it is difficult to imagine that they can be ‘lifted’ directly – there will be at least some need for them to be moderated / deliberated by either NICE’s Board, through the institutes Citizens Council or by the Appraisal Committees.

7.4 End of Life

The existing evidence in support of the current EoL criteria is mixed. For example, the study performed by researchers at Bangor University found no societal support for it while the EEPRU report found some support for survival prior to treatment. However, the authors of the EEPRU analysis clearly conclude that the concepts of EoL and BoI overlap, both in theory and in terms of empirical evidence, and state that they should not be used together to weight future QALY gains. The ToR state that the definition of BoI should ‘encompass’ EoL. However, the extent to which this achievable is unclear. For example, imagine a 70 year old person who has 2 years to live as a result of illness. In the absence of illness, they would live until 75. The BoI for this individual (75-72), ignoring QoL adjustments, is relatively small and similar to EoL. However, if the individual was 50 years of age, the overlap between EoL and BoI is much smaller.

Other recent evidence provides little support for a societal preference for additional weight for “end of life” therapies. Shah et al. [12] conducted a discrete choice experiment (DCE), funded by NICE via the DSU, to specifically investigate the EoL NICE guidance. The design was similar to the Brazier et al study, but focussed on EoL. Web-based responses from almost 4,000, UK-based individuals were received. They reported that there was no evidence that respondents on average are willing to sacrifice aggregate health gains in order to give priority to the treatment of end of life patients.

Skedgel et al. [13] also conducted a DCE amongst the Canadian general population (n=612) and health care decision makers (n=44). In contrast to NICE’s current EoL guidance, they found statistically significant preferences for treating those with *longer* life expectancies (10 yrs vs 5 yrs vs 1 month).

Abel Olsen [14] reports findings from a study of the Norwegian general population (n=503) who were asked to make a series of pair-wise choices. He reports that:

“Their preferences reveal strong support for the ‘fair innings’ argument that total lifetime inequalities should be reduced. Differences in patients’ remaining lifetime without treatment did not matter, implying little support for the ‘end of life’ argument that a short life expectancy makes patients entitled to preferential treatment.”

7.5 Valuing displaced treatments

The Reference Dataset underpins many of the DoH's proposals, particularly with regard to valuing displaced treatments. The methods of how the Dataset was compiled are complex and difficult to follow with precision. However, we believe that there are a number of important uncertainties inherent within it which will have implications for calculating the 'adjusted' thresholds; note the following is not an exhaustive list. First, the age distribution taken from the Global Burden of Disease project is not continuous – it is categorised into relatively large groups. We consider these to be crude when the extent of variation on the magnitude of WSBs within each of these groups is examined. Second, QoL estimates both in terms of burden for a disease area and gains from therapies have little linkage with individual technologies or their licensed populations. Lastly, the data set has been constructed using a data from a number of different sources. Indeed, some of it is derived from extremely crude levels of coding, such as disease categories. For example, programme budget codes (PBCs) / U codes are much broader than the 3 figure ICDs used in most places. We are unclear exactly how the two were linked together where necessary, but assume that this must lead to considerable uncertainties in the results. In summary, this means that the Dataset may not be sufficiently refined to produce a sufficiently accurate / non biased result of the nationally displaced QALY. However, it is difficult to know how inaccurate it is.

8. POSSIBLE CONFLICTS WITH OTHER ELEMENTS OF NICE'S TECHNOLOGY APPRAISALS PROGRAMME METHODS GUIDE

- **Uncertainty.** NICE's method guide [1] makes clear the importance of considering the uncertainty around an ICER in terms of structural issues (such as model design and the logic linking parameters) and parameters (in terms of how precisely they have been estimated) and how different assumptions ultimately affect the decision to recommend a treatment or not.

However, even if one were to focus purely on the BoI QALY losses (E) generated using the Reference Dataset for a specific treatment, it is unclear what these uncertainties might be (perhaps in terms of potential biases) or the error around them in a parameter sense (the values in the Reference Dataset are all expressed as point estimates; no error is included). Thus there is a danger that committees and stakeholders spend substantial time investigating the uncertainties in routine cost-effectiveness analysis, before 'adjusting' the results using a very uncertain mechanism. While a practical solution might be to let stakeholders, Assessment Groups and the Appraisal Committees undertake simple sensitivity analysis, this could be complex given the number of parameters under discussion. We note our earlier comment that the LoL and QoL losses in the absence of a treatment could be estimated using submitted economic evaluations rather than using the Reference Dataset – it is the

uncertainties in the weighting and in the valuation of the notional displaced QALY that are likely to be more problematic to estimate and to account for.

- **Uncertainty in the threshold** – Broadly speaking NICE’s Technology Appraisal Programme methods document [1] suggests that an intervention will be considered cost-effective if it is associated with an ICER under £20,000 per additional QALY, up to a value of £30,000 given additional considerations. In terms of valuing displaced treatments, it is not clear which threshold should be ‘chosen’ for adjustment, at least in the first instance, or whether these additional considerations included in the VBP proposal mean the upper value of £30,000 is now somewhat redundant. For example, one of the stated additional considerations for using the upper value is EoL. If this is now encompassed in the valuation of the benefits of the new and displaced treatments, then it would seem inappropriate to also consider a £30,000 threshold.
- **Heterogeneity** – This has already been discussed in relation to calculating ‘normal health expectancy’ in the context of multiple conditions and using the ICD codes within the Reference Dataset.
- **Time Preference** – NICE’s Reference Case clearly stipulates that future costs and benefits should be valued less (discounted) in each successive year. However, the proposed method of measuring BoI effectively appears to ignore any possible effect of time preference, meaning there is an issue of inconsistency. The DSU also notes that they are unclear how time preference relates to the derived EEPRU weights (ie whether they are effectively net of time preference or not).
- **Potential inconsistency with non-medicines within the Technology Appraisals Programme process.** While the ToR are ‘open’ with regard to applying VBP to other NICE programmes (such as Clinical Guidelines), the DSU notes the potential for inconsistencies if changes to the methods processes are not reflected in the Institute’s other programmes of work. We are also unclear as to what would happen within an individual appraisal if generic drugs are included within an appraisal scope, perhaps as comparators, which has been the case with the taxane paclitaxel. Or if a comparator treatment consisted of perhaps a surgical technique rather than a drug.
- One of the options for measuring BoI is to use the results from submitted economic evaluations rather than the DoH suggested Reference Dataset. In some appraisals, however,

more than two treatments are compared. In this instance, it is unclear which treatment the BoI (F) should relate to – that is, should the number of BoI QALYs relate to each individual drug, the next best option in terms of QALY gains (whether it represents a comparator or another intervention within a multiple technology appraisal) or a weighted average of F across the different treatment options?

- **Equality.** We believe the impact of BoI on equality is likely to be minimal compared with the incorporation of WSBs. This is because by the very nature of the proposal, groups of individuals, defined according to age, sex and current QoL will now be judged on their current ability to contribute to society in addition to their ability to benefit from treatment.

9. THE WAY FORWARD?

We do not propose to suggest a particular way forward to the Working Party in terms of implementing the BoI component of VBP. However, we have listed below a number of statements that reflect our current view of BoI that can hopefully form the basis of useful discussion points.

1. We believe there is evidence to suggest society values severity.
2. If the Working Party does not agree with this view, options, if thought to be consistent with the ToR, could be to continue with NICE's existing processes or methods with or without the existing allowance for EoL considerations.
3. There is some reason to believe that a proportionate shortfall approach might be a more appropriate definition than the 'prospective health' approach currently put forward.
4. Irrespective of whether BoI is defined as currently proposed by the DoH or in terms of proportionate shortfall, our preference would be to estimate BoI from company submissions rather than the Reference Dataset source as it is likely to be more applicable to individual appraisals, there are a number of uncertainties regarding the 'quality' of the Reference Dataset and there are less likely to be procedural complications in so much that the evidence would be fully understood by the manufacturers. We do believe this would add substantially to submission requirements since estimates of QALY loss can be estimated directly from existing models.
5. The evidence seems to suggest that BoI and EoL overlap, therefore they should not both be used as independent weighting mechanisms. However, the extent to which they overlap depends on the decision point, an individual's age and how BoI is defined.
6. The evidence suggests QoL and LoL are valued differently. The extent to which they are valued differently, however, is uncertain.
7. There are only limited options in terms of valuation. We have not seen the latest EEPRU QALY-weighting results, and they should be examined. However, given there are substantial

uncertainties around all the available weights, there is likely to be the need for either moderating the empirical estimates using a deliberative process, or basing the weights entirely on a deliberative process. It is clear that there will be a trade-off between complexity of the weighting system and ability to reflect what is known. For example, the EEPRU study shows that different weights were derived according to expected length of survival in the absence of disease and Nord emphasises the implausibility of defining BoI in a single term.

8. In this sense, a reasonable question to ask is whether the perfect should be an enemy of the merely good?
9. Although not our preferred option, an alternative to the above could be to use the EEPRU results directly since it is UK-based research that was commissioned to directly inform this aspect of BoI. In this instance, however, we would particularly question the validity and legitimacy of using only some of the results from the DCE since other attributes, such as QALY gains, were also shown to be significant and the weights are relative to each other.
10. Assuming the value of the notional displaced QALY is so to be reflected, we see no other feasible option but to use the existing Reference Dataset despite its imperfections.
11. However, there is a question regarding its ‘fitness for purpose’ given that it was not originally intended for this use. That is, whether given the existing evidence it is possible to reasonably reflect the value of a notional displaced QALY.
12. An alternative is to have a (binary) definition of severity that is so extreme that it applies only rarely and the value of displaced treatments can be ‘ignored’. However, we see no particular logic for this approach.
13. A slightly more sophisticated approach could *either* be to estimate the proportion of patients below a particular definition of severity and to weight this proportion’s benefits more highly *or* possibly estimate the proportion of the QALYs gained that qualify for a higher weighting given a definition of severity.

10. FURTHER RESEARCH

- To estimate society’s preferred definition of BoI.
- A more detailed critique of the Reference Dataset is required to assess its fitness for purpose and if necessary, further research to improve its quality.
- To investigate how countries such as Norway and the Netherlands are implementing (or not) BoI considerations.

11. REFERENCES

1. National Institute for Health and Care Excellence, *Guide to the methods of technology appraisal*, 2013.
2. Murray, C.J.L., Ezzati, M., Flaxman, A.D., et al. *GBD 2010: design, definitions, and metrics*. The Lancet, 2012. 380(9859): p. 2063-2066.
3. Claxton, K., Martin, S., Soares, M., et al., *Methods for the Estimation of the NICE Cost Effectiveness Threshold* 2013.
4. Brazier, J., Rowena, D., Whyte, S., et al., *Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing: a report of the main survey*, 2013.
5. Claxton, K., Briggs, A., Buxton, M.J., Culyer, A.J., McCabe, C., Walker, S. and Sculpher, M.J. *Value based pricing for NHS drugs: an opportunity not to be missed?* BMJ, 2008. 336(7638): p. 251-254.
6. Shah, K.K. *Severity of illness and priority setting in healthcare: A review of the literature*. Health Policy, 2009. 93(2–3): p. 77-84.
7. van de Wetering, E.J., Stolk, E.A., van Exel, N.J.A. and Brouwer, W.B.F. *Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall*. The European Journal of Health Economics, 2011. 14(1): p. 107–115.
8. Stolk, E.A., Pickee, S.J., Ament, A.H.J.A. and Busschbach, J.J.V. *Equity in health care prioritisation: An empirical inquiry into social value*. Health Policy, 2005. 74(3): p. 343-355.
9. Linley, W.G. and Hughes, D.A. *Societal views on NICE, Cancer Drugs Fund and value-based pricing criteria for prioritising medicines: a cross sectional survey of 4,118 adults in Great Britain*. Health Economics, 2013. 22(8): p. 948-964.
10. Nord, E. and Johansen, R. *Transforming EQ-5D utilities for use in cost-value analysis of health programs*. Unpublished.
11. Nord, E. and Johansen, R. *Concerns for severity in priority setting in health care: A review of trade-off data in preference studies and implications for societal willingness to pay for a QALY*. Unpublished.
12. Shah, K., Tsuchiya, A., Hole, A.R. and Wailoo, A., *Valuing health at the end of life: a stated preference discrete choice experiment*, 2012, A report by the decision support unit.
13. Skedgel, C., Wailoo, A. and Akehurst, R. *Societal preferences for the allocation of healthcare resources: a discrete choice experiment in a Canadian context*. In submission.
14. Abel Olsen, J. *Priority preferences: end of life does not matter, but total life does*. Value in Health, In press.

APPENDIX 1: CLARIFICATION ON ASPECTS OF THE THRESHOLD PROJECT

PROPOSED TO BE USED WITHIN VBP

Marta Soares and Karl Claxton, University of York. July 2013.

- The key reference is the threshold report which describes burden measures and data sources in Chapter 4, Appendix C and Addendum 1 (most up to date version of this report is available online http://www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP81_methods_estimation_NICE_costeffectiveness_threshold_revised.pdf).
- The aim of the threshold work was to develop methods to estimate the NICE cost-effectiveness threshold, making use of routinely available data. For such, programme budgeting data for the English NHS was used to estimate the relationship between changes in overall NHS expenditure and changes in mortality (econometric analysis described in Chapter 3 of the threshold report). Note that programme budgeting data which allocates the entire volume of health care expenditure to broad programme budget categories (PBCs) according to primary diagnosis. This work has extended the measure of benefit in the threshold to QALYs by estimating the quality of life (QoL) associated with additional years of life and the direct impact of health services on QoL. To do so, there was the need to describe expected burden within each PBC; data were available at ICD-10 level, however, which represents a lower level of disaggregation than PBC.
- Four aspects of this work are proposed to be further used within VBP – a brief description of each is provided below.

A. Data sources used within the threshold work to get measures of QALY burden and assumptions required for displaced QALYs

A.1 Displaced QALYs

- When NICE issues positive guidance for a new intervention which imposes additional costs on the NHS, the resources required to deliver it must be found by disinvesting from other interventions and services elsewhere. This displacement will inevitably result in health decrements for other types of individual. Thus the threshold represents the additional cost that has to be imposed on the system to forgo 1 QALY of health through displacement. The work developed estimated such threshold by evaluating the effect on health of marginal changes in the overall NHS budget based on current NHS activities. Due to data availability, the health effect analysed in the econometrics is mortality. Information on expenditure used the programme budgeting data that allocates to broad areas of illness (programme budget categories, PBCs) defined according to the primary diagnosis (using ICD10 codes) all items of NHS expenditure. Using these data, a proportional effect of spend on mortality is estimated by PBC.

- Although 3 years of mortality data are used in the analysis of each year of expenditure, these are averaged to an annual value prior to estimating outcome elasticities. Therefore, the estimated outcome elasticities represent the proportionate effect on mortality in one year due to a proportionate change in expenditure (see 3rd paragraph in page 54 and footnote 56 in report).⁴ Note that deaths averted by expenditure in one year is assumed to return the individual to the mortality risk of the general population, i.e., the years of life gained associated with each death averted are based on what would have been their life expectancy taking account of their of age and gender (using life tables for the general population) (see page 43).
- In a second stage, the proportional effect of spend estimated in the econometrics is applied to measures of QALY burden. Estimates of mortality and life year effects are thus used as 'surrogate outcomes' for a more complete measure of the health effects of a change in expenditure. This appears more plausible than assuming no effects of NHS expenditure on quality of life outcomes.
- In those PBCs where mortality effects could not be estimated, the proportional effect of changes in expenditure on QALY burden of disease is assumed to be the same as the overall proportional effect on the life year burden of disease across those PBCs where mortality effects could be estimated.
- Estimates of QALY burden were obtained from external data for ICDs (3-digit level) that compose the PBCs. QALY burden estimates are further detailed in section A.2. Burden differs across the type of diseases that make up each PBC. When using this information to estimate a cost per QALY threshold the health effects observed at PBC level must be allocated in some way to the component ICD codes, allowing calculating ICD specific QALY effects that are then summed across all the contributing ICD codes. The distribution of health effects from PBC level to ICD level is based on weighing the effects by the proportion of the total PBC population within each contributing ICD code.⁵
- Underlying estimates of the cost per QALY threshold are also estimates of the PBCs where more health is expected to be displaced, i.e. if one unit of health is to be displaced within the whole of the NHS, how much health is expected to be lost within each PBC. This has been further used by the DH and referred to as proportion of the notional displaced QALY.

⁴ This means that the health effects of changes in expenditure which reduce incidence (prevention of disease) will not be captured. (see page ix, 54 and 84 in report).

⁵ Alternatively, PBC level effects could have been allocated to ICDs based on contribution to variance in HES costs. However, HES costs are a much smaller proportion of total PBC expenditure for the 11 PBCs where a mortality effects could not be estimated (HES costs account for less than 15% of total PBC expenditure) and account for very little of the variability in PBC costs across PCTs (the contribution that variance in HES costs makes to variance in PBC expenditure in this group of PBCs is less than 8%). Therefore, allocating PBC level effects to ICDs based on contribution to variance in HES costs is less appropriate when information about QALY burden in this groups of PBCs is used to inform the estimate of the overall threshold (footnote 79 in report).

A.2 *Estimates of burden of disease*

- We aimed to estimate total QALY burden of disease for the population with disease in a particular year. This includes: i) the quality-adjusted years of life lost due to all the disease related mortality that could occur in this population over their remaining duration of disease and ii) the reduction in quality of life while alive also for their remaining disease duration.
- However, applying the estimated proportionate effects on mortality and life years to such a measure of total burden would provide an estimate of the effects of a change in expenditure, not just in one year, but in all the remaining years of disease for the population at risk in that year. To estimate the cost per QALY threshold we have adopted the conservative assumption that changes in expenditure will only have health effects in one year for the population with disease in that year. Therefore, it is not a measure of total burden that was required, but a measure of the QALY burden of disease during one year for the population with disease (prevalent and incident) in that year.
- The information used to derive QALY burden (further details ahead) includes information about the YLL and duration of disease for those incident, i.e., the measure of QALY burden is a measure of the total burden of the disease but only for the population that is incident (rather total population with disease) in one year. Assuming that incidence is stable over the disease duration this is also equivalent to the QALY burden of disease during one year for the population with disease (i.e., those that are incident and prevalent) in that year (see footnote 77 of report).
- If required, the information available does allow us to express an even fuller measure of burden of disease – the total QALY burden for the whole of the population at risk (prevalent and incident) in a year – using disease duration.
- To evaluate burden of disease for the population with disease in a particular year, a number of inputs were required. These inputs and the sources of data used are described next.
- Some of the inputs required were derived from the WHO Global Burden of Disease (GBD) study, updated in 2008 using 2004 data (see Addendum 1 in Appendix C for more details) that provide a range of summary health indicators for the UK. GBD classifies diseases by U-codes, which are groups of three digit ICD-10 codes (see Addendum 1 in Appendix C for details of how U-codes map to ICD-10 codes).
- Unlike ICD codes, U-codes do not map directly to PBCs so some ICDs in different PBCs may belong to the same U-code and therefore have the same U-code ratio. Some ICDs are not included in the U-code classification of disease. Most of these are procedural codes where we do not assign life year and QALY effects anyway (any health effects would be evident in other ICD codes), so it was not necessary to impute for them (84 out of 1562) (see footnote 46 on report). Of the others, most were associated with PBC16 with a zero outcome elasticity so did not require imputation either (186 out of 1562). Imputation across the ICDs within the PBC was required for the remaining (482 out of 1562). Eighty eight of these cannot be mapped into U-codes. The remaining

394 were associated with U-codes where mortality and YLL were zero. In both these cases, values were imputed across the ICDs within the PBC (see footnote 71 in report).

- Age and gender distribution of the at risk population: Each U-code was sub-divided by disease sequelae which represent disease sub-categories of each U-code (Addendum C1 to Appendix C). As an individual may be represented in multiple sequela in a single U-code to avoid double counting in the event of multiple sequela in a given U-code our analysis uses prevalence estimates based on the sequela with the largest prevalent population. The age and gender distributions used here are based on those reported within GBD for such prevalent population. GBD uses 8 age groups – where needed, results have been averaged over the age groups using the midpoint age of each range.
- Norm QoL scores and disease decrements: There is good evidence that, on average, the general population is not in full health, i.e. the quality of life score associated with the health states experienced by the general population are less than 1, decline with age and differ by gender. These quality of life ‘norms’ for the general population by age and gender were based on an analysis of data from the Health Survey for England (HSE) (see Addendum C1 for a description on HSE data).

We were also able to evaluate the expected quality of life decrement associated with each ICD code (i.e. for diseased patients) using existing datasets. The Health Outcome Data Repository (HODaR) provides over 30,000 observations of EQ-5D measures of quality of life by ICD code and the age and gender of the patients in the sample (see Addendum C1). Although this is a rich UK data set, there were a limited number of observations for some of the less common ICD codes. For this reason HODaR was supplemented with information from the Medical Expenditure Panel Survey (MEPS) which also provides EQ-5D by ICD and reports the average age of respondents (see Addendum C1). These data provided a means of estimating the quality of life associated with each ICD code at the average age of respondents in the pooled sample (datasets were pooled considering the number of patients from each contributing to estimates, i.e. a weighted average).

The average quality of life scores across the ICDs which contribute to each U-code and the average age and gender of respondents from HODaR and MEPS were used to calculate a disease decrement, based on quality of life norms from the general population. These disease decrements can then be applied to the age and gender distribution of each U-code. Note that this decrement is fixed (subtractive), rather than proportionate (see footnote 61 in report).

- Duration of disease: From GBD we used duration of disease for those incident to a U-code. By assuming constant incidence, the average remaining duration of disease for those at risk was then calculated.
- Years of life lost (YLL): GBD also provides information on deaths and YLL by age and gender; however, in using absolute measures of burden it becomes more important to examine and then

adjust for any inconsistency between information about YLL and size of the incident population from GBD (which is available by U-codes and can be mapped to ICDs), and the information about net YLL and observed deaths for each PBC based on ONS data as described in Section 4.2.3.⁶

- The information used to derive QALY burden is based on information about patients incident to a U-code, i.e., the measure of QALY burden is a measure of the total burden of the disease but only for the population that is incident (rather total population with disease) in one year. Assuming that incidence is stable over the disease duration this is also equivalent to the QALY burden of disease during one year for the population with disease (i.e., those that are incident and prevalent) in that year.⁷

B. the additional assumptions required if weight QALY effects according to burden

- The threshold project provides information on the proportion of a QALY displaced in each ICD. Such QALY effects may be weighted using burden of illness weights and a mechanism for weighting that we will here assume to be known, to generate a ‘multiplier’ for the displaced activities.
- If length or quality effects do not need to be distinguished, then no further assumptions are required other than those used to generate proportion of displaced QALYs across ICDs and burden estimates (mentioned in A.1 and A.2).
- If length and ‘pure’ quality effects need to be distinguished, a further assumption is required on how a proportionate effect on total burden affect each component (such assumption was not needed for the threshold project). We have assumed that these components are affected proportionally to their contribution to overall burden (thus, for example, any effects on diseases that mainly compromise LoL will mainly occur over this component of burden)

⁶ There are a number of reasons for potential inconsistencies: i) GBD is based on earlier years of mortality data; ii) the imprecision of mapping from U-codes to PBC via ICD codes; and iii) the YLL reported in GBD are calculated in the same way as published NHS IC estimates (see Section 4.2.2 and 4.2.3) and will tend to overestimate the net YLL (see Table 52 in Appendix C). The YLL by U-code, reported in GBD, that are mapped to ICDs are adjusted by these proportionate differences to ensure that the YLLs associated with all contributing ICD codes are consistent with (do not over estimate) the net YLL for the PBC as a whole. However, due to the earlier years of data and imprecision in mapping from U-codes to ICDs there might also be some inconsistency in estimates of the total incidence of disease for a PBC. Insofar as disease related mortality risk is stable, the same number of deaths should be observed in GBD and ONS data for the same at risk population. The PBC deaths recorded in GBD and those observed in ONS data (see Table 52 in Appendix C) are similar but nonetheless the proportionate difference is used to adjust the scale of quality of life burden while alive based on GBD information (equivalent to adjusting estimates of incidence). Notable exceptions are PBC1 and PBC18+19 where the discrepancies are due to imperfect mapping from U-code to PBC via ICD codes.

⁷ So long as estimates of the quality of life decrement of disease from HODaR and MEPS are representative of average effects across those earlier (incident) and later (prevalent) in their disease duration an assumption of constant quality of life decrement with respect to disease duration is not required.

C. When WSB are function of length or quality of life effects as well as age, gender and ICD code etc.

- Again, the threshold project provides information on the proportion of a QALY displaced in each ICD. Such QALY effects may be weighted using WSBs and a mechanism for weighting that we will here assume to be known, to generate a ‘multiplier’ for the displaced activities.
- If such process only depends on burden of illness, QoL score of diseased (starting QoL score), age, gender and ICD code (3-digit or any higher level of aggregation), then no further assumptions are required other than those used to generate proportion of displaced QALYs across ICDs, i.e. burden of illness, age and gender (mentioned in A.1 and A.2).
- However, if WSB depends on QoL score ‘after displacement’ (finishing QoL score), further assumptions are required. Firstly, for each QALY displaced in an ICD, QALYs associated with pure QALY effects need to be isolated from length of life QALY effects (see point B for the extra assumptions this requires). From total QALYs associated with pure QoL effects, and by assuming a constant effect over QoL over time, the duration of effect can be used to evaluate the QoL score ‘after displacement’ (finishing QoL).

D. Additional assumptions that might be required if these estimates are used as defaults on the benefits rather than displaced side.

- The information described above can provide relevant evidence for calculating WSB and burden weights in the benefits side of health technology appraisals. Specifically, information on the age and gender distribution could be used to describe the eligible population in an appraisal. Also, information available on expected QoL decrement due to disease could also be relevant.
- We feel that it would be consistent to use such information on both displaced side and benefits side. However, this information is unlikely to be regarded as plausible for appraisals of particular technologies, namely due to the following:
- The level of 3 -digit ICD, for which we have information on ‘at risk’ population from GBD (prevalent and incident), may be unrepresentative of the population of interest for appraisal.
- The information from GBD relates to the year 2004, and may be judged unrepresentative if incidence, duration of disease or mortality changed over time (possible, for example, if NHS care changed). The same argument can be applied to evidence on QoL score.
- To apply the information available from HoDAR and MEPS on QoL decrements one would need to assume QoL to be constant over the duration of disease. However, more detailed and accurate information on how QoL evolves is commonly available within the appraisals.
- The evidence on age and gender distribution of the at risk population in GBD relates to a point in time (one particular year) and reflects the joint effect of incidence, duration of disease and mortality under NHS care up do that point in time. GBD evidence does not tell us directly how the

age and gender distributions of a cohort of patients changes over time by mortality, which may be important to evaluate WSB and burden weights.

- Also, GBD (or HoDAR and MEPS) cannot inform of the effect of a ‘new’ treatment on the age and gender distribution over time.