

Methods for estimation of the NICE cost-effectiveness threshold

Summary of research based on proposal to MRC

Importance

A comparison of the incremental cost effectiveness ratio (ICER) of a new technology with a cost-effectiveness threshold is not the only consideration when the National Institute for Health and Clinical Excellence (NICE) and its advisory committees issues guidance. But is an important one: it allows an assessment of whether the health expected to be gained from the use of a technology exceeds the health expected to be forgone elsewhere as other NHS activities are displaced. For this reason a comparison of the ICER of a technology to a threshold range is a critical part of the reference case in the NICE Guide to Methods of Appraisal¹ and is often taken to be the starting point for deliberations about other consideration including judgements of social value. Therefore, the value of the threshold or the range of values used is critical to the assessment of whether technologies can be regarded as cost-effective with implication for NHS patients, local NHS decision makers, the Department of Health, HM Treasury, manufacturers (pharmaceuticals and devices) as well as NICE itself.

i) What is the cost-effectiveness threshold? – In principle, the cost effectiveness threshold is an estimate of health forgone as other NHS activities are displaced to accommodate the additional costs of those technologies recommended in NICE guidance.² A national decision-making body like NICE needs an estimate of what is likely to be forgone on average across the NHS as we currently find it. Of course, this will change as circumstances and the NHS changes; tending to rise with increases in budget and health care costs but tending to fall with increases in the productivity of health technologies and the efficiency of the NHS in general - including better local commissioning decisions.³ A body like NICE cannot and does not necessarily need to know what *specific* services and treatments will be displaced in particular localities or who will actually forgo health. What is required, however, is an accountable and empirically-based assessment of the health that is likely to be forgone on average across the NHS.

ii) What are current estimates based on? – Currently NICE uses a threshold range of £20,000 to £30,000 per quality adjusted life year (QALY) gained, where additional considerations are required towards the upper bound. The empirical basis of this range of values is very limited. At best it represents an informal assessment of the health gained by some of the least productive (in health outcome terms) of the activities currently undertaken by the NHS. Unsurprisingly, it does, to a certain extent, represent the implied values from past NICE decisions. It is widely recognised, by NICE⁴ and the House of Commons (HoC) Health Committee⁵ among many others,⁶ that the current range ought to be more firmly based on empirical analysis. The HoC Health Committee highlighted particular concerns that the additional costs of NICE guidance imposed on local NHS commissioners might be causing the displacement of more valuable health care.⁵ On the other hand, manufacturers and others have argued that the threshold range is too low, restricting market access, prices and revenue and ought to be based on how much individuals are willing to pay for improvements in health.⁷ In 2009 NICE convened a workshop to discuss what the threshold ought to represent and how it might be more securely estimated.⁸ Most of the applicants contributed to that workshop and it has informed the plans for research set out below.

iii) What is needed? – Explicit scientific methods for estimation are required which will provide accountability so that estimates can be scrutinised by a range of stakeholders. Since estimates of the threshold will need to be periodically revised, methods which make best use of routinely available NHS data are needed. As well as accountability, this will also provide more predictability in likely changes to the threshold for the investment decisions of technology manufacturers. Providing more secure estimates of the threshold will necessarily require application of specialist expertise and the development and application of sometimes sophisticated methods. Nevertheless these must be communicated effectively to stakeholders to ensure transparency. Suitable methods should provide estimates relevant to NICE - that is, relevant across the NHS; they should capture the effects of health care on both length and quality of life, offer the opportunity to estimate changes in the threshold over time and indicate the impact of 'non marginal' changes which have a significant budget impact on the NHS.

Research Plans

The experience of attempts to look in some detail at the decision making processes and outcomes at a local level have demonstrated the complexity, variability and difficulty of estimating the cost effectiveness thresholds implicit in local NHS decision making.^{10,15} Although such studies provide

valuable insights into the nature of local decision making, they are unlikely to meet the needs set out in section 2 because: i) they do not readily provide the national picture NICE requires; ii) it is not clear how they could be used to estimate changes over time, the effects of non-marginal changes or add to predictability; and iii) such detailed studies are very much bespoke and specific to time and place, so the more granular view they provide cannot feasibly be routinely collected. In addition, it may not be necessary to know precisely which health technologies are displaced, for which patients or why. What is required is an estimate of the health that is likely to be forgone *on average* across the NHS, ideally based on routinely available data. Therefore, this research will focus on complementary methods which can make best use of those data that are already available, where there are already plans to make data available or where additional data could feasibly be made available at reasonable cost. The research plans fall into the following 4 complementary areas of activity, all of which will be evaluated at a user impact workshop.

3.3.1 Review of principles, methods and estimates of the threshold

The literature which considers the cost-effectiveness threshold has grown over recent years.³ However, most is rather discursive,^{7,16} some more analytic,^{2,9} but only a few attempts to offer methods for empirical estimation.^{10,12-14} This diverse literature (including policy documents) needs to be thoroughly reviewed and pulled together in a clear, comprehensive and structured way. This will: i) clarify the terms of the debate and establish a common understanding of concepts and the types of estimates that are required; ii) review how different approaches to estimation seek to meet the needs outlined in 2); and iii) outline what is missing from current estimates and how methods might be most usefully developed. This review will be written for a wide policy audience, but will also be comprehensive; covering all the issues raised in recent debates and explain alternative methods of estimation in an accessible way. The review may inform the basis of an agreed framework with all relevant bodies (NICE, DoH, HoC Health Committee, HM Treasury etc). It might also inform the remit of the type of independent body suggested by HoC Health Committee⁵ or the periodic reviews suggested by NICE.⁸

Analysis of programme budget data

Since 2003 data on expenditure on health care across 23 programmes of care have been prepared by each Primary Care Trust (PCT) in the English NHS. These programme budgeting (PB) data seek to allocate exhaustively to disease categories (via ICD10 codes) all items of NHS expenditure, including expenditure on inpatient care, outpatient care, community care, primary care and pharmaceuticals and devices. It serves a number of purposes, notably to assist in the local planning of healthcare. But its crucial merit for this study is that it opens up the possibility of examining the relationship between local spending and associated disease-specific outcomes.

Previous work by some of the applicants has demonstrated the potential value of programme budgeting data in estimating the link between expenditure on a programme of care and the health outcomes achieved, in the form of disease-specific mortality routinely available from the National Centre for Health outcomes Development. In each programme changes in mortality associated with changes in expenditure are transformed into life years, providing estimates of the marginal cost per life-year gained on average across the NHS. This work has focused largely on spending and outcomes in two of the largest programmes of healthcare: circulatory disease and cancer¹², but has also informed the link across other programme categories^{13,14}. Estimates of the cost per life year gained for 2006/07 are £15,387 for cancer; £9,974 for circulation problems; £5,425 for respiratory problems; £21,538 for gastro-intestinal problems; and £26,428 for diabetes. These estimates are based on a straightforward, though carefully constructed, theoretical model of health production which informs the specification and estimation of a system of equations (issues of endogeneity are dealt with by identifying and testing suitable instruments). In doing so, they account for variation in the clinical needs of the local population relevant to the programme of care and broader local environmental factors relevant to the costs of care and outcomes. In principle, this approach, based on routine data, estimates the type of cost-effectiveness threshold required by NICE: the 'average' marginal elasticity of spending with respect to income amongst the PCTs. However, the methods of analysis need to be developed in a number of important respects; not least to express outcome in terms of quality as well as length of life (see 3.3.3). These are outlined below.

i) What is the overall threshold for the NHS? - The overall threshold for the NHS will depend on the programmes of care where disinvestment takes place. Hitherto, each programme of care has been estimated separately so it is not clear how expenditure on particular programmes changes with the overall budget, e.g., does disinvestment tend to fall on respiratory care or diabetes? In principle, spending on programmes is linked by a system of equations, and

we will seek to model the programme expenditure equations as a set of linked simultaneous equations, reflecting the potential for interactions between spending on the different programmes, brought about by the need for PCTs to operate within a fixed overall budget. This will offer an opportunity to study the 'budget elasticity of expenditure' in each programme of care. It is then feasible to derive estimates of the impact of marginal increases (or decreases) in overall PCT budgets on spending in each of the programme categories. As well as indicating budgetary influences on programme spending these can then be linked to changes outcomes by programme. This can provide an estimate of the cost per life year gained on average across the NHS, for marginal changes in budget. This type of estimate of the threshold will take account of how such budgetary changes translate through local decisions into changes in expenditure on programmes of care and then to health outcomes. In addition, knowledge of budget elasticities of expenditure across the different programmes of care, coupled with the estimated relationship between expenditure and health outcome, will indicate how expenditure changes in one programme impacts on health outcomes in other programmes of care, providing more secure estimates of the relationship between overall expenditure and overall health outcome.

ii) What is the likely impact of non-marginal changes? - To the extent that data permit, we shall seek to study year-on-year changes in spending, as well as a cross section of spending decisions. Changes in budgets are in practice incremental, and it may be the case that the elasticities of programme expenditure in times of budgetary increase (when new initiatives are introduced) are not the same as in times of budgetary decrease (when the focus is on disinvestment). They may also vary depending on the current level of expenditure (relative to need) on a specific programme of care. In general, elasticity might be expected to increase as spending increases, but this can be tested. This offers the opportunity to explore the possible effect of non-marginal changes on programme expenditure and possibly outcome, providing estimates of the threshold for a range of budget impacts. This type of analysis would provide some guidance to NICE on when a decision might regarding a new technology have such a significant impact on the NHS budget that there will be significant reallocations of expenditure between programmes, with more valuable health care forgone so that a lower estimate of the threshold might be appropriate. It might also suggest when a series of apparently marginal changes (mandatory NICE guidance) will start to have non-marginal effects on the NHS.

iii) How do estimates change over time? - We will investigate whether successive years of PB data can be used to form a panel dataset. This might be limited by changes to PCT boundaries but, should it prove feasible, a robust panel will allow an assessment of the stability of our estimates. It will also allow: an investigation of how elasticities and estimates of the threshold change over time; an assessment of the feasibility of periodically re-estimating the threshold based on these types of data; and an exploration of the possibility of making predictions based on overall budget forecasts. Previous analysis of PB data assumed a quasi long-run equilibrium so that health outcomes could be contemporaneously linked to expenditure. This is likely to be more tenable in some programmes of care than others. However, should it prove feasible to construct a robust panel, we will investigate empirically the appropriateness of the assumption of equilibrium and whether the relationship between programme expenditures and health outcomes is better represented by data lags and how a suitable lag structure might differ between programmes. We will also investigate the potential of using other data sources to complement the programme budget data. For example, Hospital Episode Statistics, while restricted to secondary care, may allow analyses at a more granular level than the PB data alone. This may include small area data on health care needs and supply which, together with practice Quality and Outcomes Framework (QOF) data, may provide a better means of adjusting for variation across practices within and between PCTs.

Overall the analysis will yield a set of budget elasticities of expenditure, disaggregated by programme of care and (if feasible) by level of spending with appropriate links between programme expenditure and outcomes over time. An important policy question is, then, the impact of a policy 'shock' on spending patterns, over time and between programmes. In principle, the study should yield information needed to examine the dynamic impact of such a shock, using simulation methods such as system dynamics.¹⁷ The feasibility and specification of this approach will depend on the information secured in the earlier stages, but the study will seek to develop such a model, to the extent that data permit.

Evidence of quality of life

The link between variations in budget, changes in expenditure on programmes of care and health outcomes, in the form of disease-specific mortality described in 3.3.2, needs to be extended so that outcome can be expressed in terms of quality of life. The previous analysis of PB data made

this link by assigning quality of life estimates (by ICD10 from the Health Outcomes Data Repository (HODaR)) to the change in life years estimated from disease specific mortality within each programme. However, the reported costs per QALY gained for each programme did not capture improvements in quality of life independent of effects on mortality, tending to overestimate the programme-specific thresholds, particularly in those programmes where expenditure tends to be associated with improvements in quality rather than mortality. A cost-effectiveness threshold for NICE needs to be expressed as the cost per QALY gained. Therefore a more complete picture of the quality of life outputs from the NHS spending in 3.3.2 is needed. Additional work is required to evaluate complementary sources of evidence and methods of analysis which will allow them to be combined with the results from the econometric analysis outlined in 3.3.2. There are two areas in particular where additional work is required:

i) Weighting improvements in length of life - The estimated gains in life expectancy associated with reductions in disease-related mortality need to be weighted to reflect the health-related quality of life (HR-QoL) of the additional years of life. As each year of life gained is not experienced in full health, costs per life-year gained will tend to underestimate the cost per QALY gained. Also, the HR-QoL of additional years is likely to differ between programmes. Therefore it is important incorporate evidence of HR-QoL to obtain a more complete picture of the relationship between expenditure and health outcome. This will be addressed two ways. Firstly, by reviewing, in each programme and associated ICD10 chapters, published quality of life evidence, exploiting existing databases of quality of life studies, e.g., the TUFTS Cost-Effectiveness Analysis Registry, EuroQol group. It is anticipated that this will reveal considerable heterogeneity between, but also within, programmes. These variations might suggest the need to complement aggregate data at programme and ICD chapter level (see 3.3.4) with other evidence. Any systematic differences and variations will provide a basis for a sensitivity analysis around the quality of life weightings applied to programmes. Secondly, other sources of data currently available (see HODaR below) or likely to become available in the future (see PROMs below) will be reviewed for suitability and quality. Where possible, the impact of using alternative sources of HR-QoL evidence on estimates of the threshold will be demonstrated through sensitivity analysis.

ii) Capturing improvements in quality of life - NHS resources are often spent on services which do not have the aim of reducing mortality, but rather of improving quality of life. Failing to account for this type of output will underestimate health outcomes and overestimate the cost per QALY gained, with the scale of overestimation differing across programmes. This will be addressed in two ways. Firstly, the review of published HR-QoL evidence described above will distinguish QALY gains arising from reductions in mortality and improvements in quality alone. We will examine whether there are any systematic differences between programmes in the proportion of health gain arising from improved length or quality of life. Again, we expect to find heterogeneity within and between programmes which might suggest that complementing aggregate data with other evidence of displacement could be useful (see 3.3.4).

Secondly, we will investigate the use of disease specific data sets, which report HR-QoL and may supplement the type of analysis in 3.3.2. For example, the Patient Reported Outcome Measures (PROMs) initiative was introduced in April 2009. The pilot for PROMs provided condition-specific and generic health outcomes data before and after surgery for four elective procedures, which allows analyses of the relationship between variation in spending and improvements in QALYs^{11,18} by procedure. The PROMs project is to be rolled out over a range of chronic conditions in the near future. However, it is unlikely that these data will be available within the timeframe for this project. Nevertheless, establishing the method of analysis that will be needed to make best use of these data will be useful, so they can complement the analysis in 3.3.2 when they become routinely available. We will also investigate the use of HODaR which collects single observations of the EQ-5D profile from patients 6 weeks following a range of procedures and services delivered in secondary care. These data record ICD10 chapter and may provide a means of estimating the HR-QoL of patients for a range of interventions within and between programmes. Unlike PROMS there are no repeated measures with HODaR, so estimating *gains* in HR-QoL will not be possible. Nevertheless, it may provide a means of establishing post-treatment QALYs within programmes and, together with evidence from the review, indicate the variation within and between programmes. We are aware of, and will investigate further, the availability and potential usefulness of other NHS datasets collecting only disease-specific measures. For example, some data are routinely collected in mental health, using instruments such as HoNOS and COR-OM. However, there are concerns about data quality and the difficulty of translating these measures into QALYs.¹⁹

The analysis of elasticities described in 3.3.2 will be used to identify those programmes which are a particular priority for review and analysis of quality of life estimates. For example, it

may be that only a few ICD chapters account for most of the changes in programme expenditure due to marginal changes in budget. By prioritising in this way we will be able to focus the search and analysis of quality of life estimates on those programmes which are most critical to the cost-effectiveness threshold. Based on estimates over time, this type of approach will also be used to identify priorities for future routine data collection in those programmes which are most critical to estimates of the threshold.

Evidence of investment and disinvestment

Although the analysis of PB data will provide estimates of how programme expenditure responds to marginal changes in overall spending it cannot, by itself, provide details of how changes in expenditure on a particular programme are allocated within the programme, e.g., the services, treatments and procedures invested in or disinvested. As discussed above, diseases-specific mortality is only one aspect of outcome, but there is likely to be significant heterogeneity within and between ICD chapters in quality of life. Therefore, it would be valuable to have more detailed evidence about the types of investments and disinvestments made within programmes and ICD chapters. Other sources of evidence would enable use of estimates quality of life which more closely matched the types of investment and disinvestment which lie behind the more aggregate changes in programme expenditure and complement the econometric analysis of PB data in 3.3.2. A complete and detailed picture of all investment and disinvestments across the NHS is not feasible on a routine basis nor would it be necessary. Our focus will be on what can be gained from other routinely collected data at a local level, and what additional evidence would be most useful and could be gathered at reasonable cost. We will also explore how and whether such evidence might contribute directly to the quantitative estimates from 3.3.2 and 3.3.3, or provide useful contextual information and help with a qualitative assessment of the considerations that ought to be applied when interpreting estimates of the threshold.

We will be working with the support and guidance of Professor David Parkin, Chief Economist at South East Coast SHA to identify and evaluate potentially useful sources of evidence that are already, or planned to be, collected at a local level (PCT and Trusts). For example, from March 2009, PCTs have been required to provide documents that detail their policies on services that they will not normally fund. Working with SHAs, we will collate these lists across the NHS. Initial communications from one SHA indicate their PCT lists contain nearly 100 such items. We will also examine the 'pledges' for new services or service delivery routinely reported in the Operating Frameworks of SHAs and PCTs. We will link all these items to programmes in 3.3.2 and to the review of quality of life in 3.3.3, whilst anticipating that some items may require bespoke review of published literature. We will explore what impact this more granular information on investment and disinvestment within programmes has on estimates of programme-specific expenditure per QALY gained and estimates of the overall threshold. However, we recognise that these data are unlikely to report all forms of investment and disinvestment (e.g., scale of services and eligibility for treatment). For this reason, we will recruit and work with NHS organisations within the South East Coast SHA to design and pilot additional data collection to supplement that available from the sources described above. We will prioritise any additional data collection based on the analysis in 3.3.2 and 3.3.3. For example, we will focus the pilot of additional data collection on those programmes which account for most of the changes in expenditure (high elasticities) due to changes in overall budget. It is estimating the QALY impact of these changes which contributes most to the cost-effectiveness threshold. While evidence from this pilot data collection may not be representative of the NHS as a whole, it will show whether additional data would be a useful complement to national PB data and routinely collected quality of life measures. It will also identify whether additional data collection is likely to be feasible on a routine and representative basis. The analysis of the PB data will help identify what might constitute a representative sample of PCTs for these purposes.

User impact assessment

We plan to present and assess the impact of this research through a workshop which will involve a range of key stakeholders. We hope that the workshop will be under the auspices of NICE and would be a full day, adopting a similar format to the recent methods workshops which informed the revision of the NICE Methods guide. All material will be pre-circulated, and will clearly pose the questions to be addressed. The presentations of the three aspects of the work described above will be followed by group discussion with facilitators and note takers with feed back on the day. The results will be circulated to participants for comment before a final report is produced. We will consult our advisory group and the Institute about possible workshop participants. However, we anticipate that they will include representation from key stakeholders including:

NICE; PCTs and other local NHS decision makers; Department of Health (including those with responsibility for commissioning and pharmaceutical pricing); HoC Health Committee; HM Treasury; National Audit Office; industry (pharmaceutical and medical devices); and academics from the relevant research communities. The primary output will be a series of recommended options that NICE may choose to take forward for public consultation.

References

1. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. NICE: London, 2008.
2. Culyer AJ, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, Sculpher M, Brazier J. Searching for a threshold, not setting one: the role of the National Institute of Health and Clinical Excellence. *Journal of Health Services Research and Policy*, 2007; 12(1): 56-58.
3. McCabe C., Claxton K. and Culyer AJ. The NICE cost-effectiveness threshold: what it is and what it means. *Pharmacoeconomics*, 2008; 26(9): 733-744.
4. House of Commons Health Committee. National Institute for Health and Clinical Excellence: NICE response to the first report of session 2007-2008. HC550. London: Stationery Office, 2008.
5. House of Commons Health Committee. National Institute for Health and Clinical Excellence. First report of the Health Committee 2007-2008. HC27-I. London: Stationery Office, 2008.
6. The government's response to the Health Select Committee's first report of session 2007-08 on the National Institute for Health and Clinical Excellence. Cm7331. London: Stationery Office, 2008.
7. Towse A. Should NICE's threshold for cost per QALY be raised? Yes. *BMJ*, 2009; 338: 268-269
8. National Institute for Health and Clinical Excellence. Threshold workshop: report of a technical meeting organised by NICE. NICE: London, 2008.
9. Devlin N, Parkin D. (2004) Does NICE have a cost effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics* 13(5): 437-52.
10. Appleby J, Devlin N, Parkin D, Chalkidou K, Buxton M. Searching for cost effectiveness thresholds in the NHS. *Health Policy* (forthcoming 2009).
11. Devlin N, Parkin D, Browne J. (2009) Using the EQ-5D as a performance measurement tool in the NHS. *Economics Discussion Paper 09/03*, Economics Department, City University.
12. Martin, S., Rice, N. and Smith, P. (2008a), "Does health care spending improve health outcomes? Evidence from English programme budgeting data", *Journal of Health Economics* 27, 826-842.
13. Martin, S., Rice, N., and Smith, P. (2008b), "Further evidence on the link between healthcare spending and health outcomes in England", *The Health Foundation*, London, May, pp61.
14. Martin, S., Rice, N., and Smith, P. (2009), "The link between health spending and health outcomes for the new English primary care trusts", *The Health Foundation*, London, June, pp64
15. Williams I, McIver S, Moore D, Bryan S. (2008) The use of economic evaluations in NHS decision-making: a review and empirical investigation. *Health Technol Assess.* 2008 Apr;12(7):iii, ix-x, 1-175.
16. Raftery J. Should NICE's threshold for cost per QALY be raised? No. *BMJ*, 2009; 338: 268-269.
17. Smith, P. and van Ackere, A. (2002), "A note on the integration of system dynamics and economic models", *Journal of Economic Dynamics and Control*, 26(1), 1-10.
18. Browne, J., Jamieson, L., Lawsey, J., van der Meulen, J., Black, N., Cairns, J., Lamping, D., Smith, S., Copley, L., Horrockes, J. Patient Reported Outcome measures (PROMs) in elective surgery. Report to the Department of Health, 2007.
19. Jacobs R. Investigating patient outcome measures in mental health . Background paper, OHE Commission on NHS outcomes, performance and productivity. OHE London: OHE, 2007