Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold

Karl Claxton,[1,2] Steve Martin,[2] Marta Soares,[1] Nigel Rice,[1,2] Eldon Spackman,[1] Sebastian Hinde,[1] Nancy Devlin,[3] Peter C Smith[4] and Mark Sculpher[1]*

[1]Centre for Health Economics, University of York, York, UK
[2]Department of Economics and Related Studies, University of York, York, UK
[3]Office of Health Economics, London, UK
[4]Imperial College Business School and Centre for Health Policy, Imperial College London, London, UK

*Corresponding author

Declared competing interests of authors: Karl Claxton, Nigel Rice, Mark Sculpher, Nancy Devlin, Marta Soares and Eldon Spackman have undertaken consultancy/research for pharmaceutical manufacturers and/or other commercial life sciences companies that may have an interest in this research. Peter C Smith is a member of the NHS Co-operation and Competition Panel, and the policy board of the Office of Health Economics. Nancy Devlin is employed by the Office of Health Economics, which receives funding from the Association of the British Pharmaceutical Industry.

Published February 2015
DOI: 10.3310/hta19140

Scientific summary

Methods for the estimation of the NICE cost-effectiveness threshold

Health Technology Assessment 2015; Vol. 19: No. 14
DOI: 10.3310/hta19140

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Introduction

The National Institute for Health and Care Excellence's (NICE) comparison of the incremental cost-effectiveness ratio of a new technology, which is more costly than existing alternatives, with the cost-effectiveness threshold is important in assessing whether or not the health expected to be gained from its use exceeds the health expected to be forgone elsewhere as other NHS activities are displaced (i.e. whether or not the new technology is cost-effective).

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 individuals. Thus, the threshold represents the additional cost that has to be imposed on the system to forgo 1 quality-adjusted life-year (QALY) of health through displacement.

Currently NICE uses a threshold range of £20,000–30,000 QALY gained, and this has remained the case in the NICE methods guidance since 2004. There have been a number of calls for further research on the value of the threshold.

This report details a 2-year project, funded by the National Institute for Health Research (NIHR) and Medical Research Council (MRC) Methodology Research Programme, to develop methods to estimate the NICE cost-effectiveness threshold.

The NICE remit implies a series of characteristics for any empirical research on the threshold:

- Reflect the expected health effects [in terms of length and quality-of-life (QoL)] of NICE guidance through the displacement decisions taken across the NHS rather than what specific services are (or could have been) displaced.
- Facilitate regular updates, based on routinely available data, to reflect NHS changes such as real overall expenditure and productivity. This would encourage accountability through scrutiny by stakeholders and provide predictability for technology manufacturers' investment decisions.
- The nature of service displacement and the magnitude of the health forgone will depend on the scale of the budget impact which should, ideally, be reflected in the value of the threshold.
- Methods should recognise the inevitable uncertainty relating to the evidence currently available for the threshold and reflect its implications for policy.

Study methods

The aim was to develop methods to estimate the NICE cost-effectiveness threshold making use of routinely available data. Objectives were:

i. informed by relevant literature, to provide a conceptual framework to define the threshold and the basis of its estimation
ii. using programme budgeting (PB) data for the English NHS, to estimate the relationship between changes in overall NHS expenditure and changes in mortality

iii. to extend the measure of benefit in the threshold to QALYs by estimating the QoL associated with additional years of life and the direct impact of health services on QoL
iv. to present the best estimate of the cost-effectiveness threshold for policy purposes.

Earlier econometric analysis estimated the relationship between differences in primary care trust (PCT) spending and associated disease-specific mortality. Expenditure came from PB data which allocates the entire volume of health-care expenditure to broad programme budget categories (PBCs) according to primary diagnosis.

This research extended this in several ways including estimating the impact of marginal increases or decreases in overall NHS expenditure on spending in each of the 23 PBCs. These were linked to changes in mortality outcomes by PBC across 11 PBCs.

The results of the econometric analysis were translated into broader effects in terms of QALYs. The first stage linked estimated effects on mortality to life-years taking into account the 'counterfactual' deaths that would have occurred if the population in a given PBC faced the same mortality risks as the general population. The second stage accounted for the health (QALY) effects of changes in mortality due to changes in expenditure reflecting how QoL differs by age and gender. The third stage incorporated those effects on health not directly associated with mortality and life-year effects (i.e. the 'pure' QoL effects) to estimate an overall cost per QALY threshold. The approach uses the estimates of mortality and life-year effects 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 QoL outcomes.

The estimated proportional effect on the mortality and life-year burden of disease is applied to measures of QALY burden. Applying a proportionate effect to measures of QALY burden of disease is equivalent to assuming that any estimated effects on life-years are lived at QoL that reflects a proportionate improvement to the QoL with disease. It also allows QoL effects of changes in expenditure to be included, also based on proportionate improvement in the QoL with disease. 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.

The methods planned for the study included a consideration of local data, collected routinely by PCTs, on the types of intervention in which local decision-makers were investing and disinvesting. The aim was to inform the link between the effects of expenditure changes on mortality and impacts on broader health in terms of QALYs. These data may have indicated the types of interventions and services, within a given PBC, on which investment and disinvestment were taking place. Using targeted literature reviews, estimates of QoL for those activities may have been identified. However, it was established that there were limited data available at a local level to facilitate this type of analysis, so other data sources were used for this purpose.

Central or 'best' estimate of the threshold

The most relevant threshold is estimated using the latest available data (2008 expenditure, 2008–10 mortality). The central or 'best' threshold is estimated to be £12,936 per QALY.

Which programme budget categories have the greatest influence on the overall threshold?

Although the 11 PBCs where a mortality effect of changes in expenditure could be estimated only account for 50% of the change in overall expenditure, they account for 78% of the overall health effects. The other 12 PBCs, where mortality effects could not be estimated, account for an equal part of a change in overall expenditure (50%) but only 22% of the overall health effects (i.e. the cost per QALY estimates associated with a change in expenditure in these PBCs are, in general, much higher).

Insofar as investment and disinvestment opportunities in these PBCs might have been more valuable (offered greater improvement in QoL) than suggested by the implied PBC thresholds, the overall QALY effects will tend to be underestimated and the overall cost per QALY threshold will be overestimated.

The overall threshold of £12,936 may be conservative (i.e. could be overestimated) with respect to health effects in PBC5 (mental health disorders), which accounts for a large proportion of the change in overall expenditure (18%) and contributes most to the overall health effects (12%) compared with other PBCs. The cost per QALY associated with this PBC is based on an extrapolation rather than observations of the direct impact of changes in expenditure on QoL. Available evidence suggests that the investment and disinvestment opportunities in mental health may have been more valuable than its implied cost per QALY.

How uncertain are the estimates and what are the implications?

Simulation methods were used to reflect the combined uncertainty in the various estimates from the econometric analysis. This indicated that the probability that the overall threshold is < £20,000 per QALY is 0.89 and the probability that it is < £30,000 per QALY is 0.97.

As the consequences of overestimating the threshold are more serious than underestimating it in terms of population health, a *policy threshold* will be lower than the mean of the cost per QALY threshold (i.e. lower than £12,936) to compensate for the more serious consequences of overestimating the 'true' value.

There were other ('structural') sources of uncertainty associated with the estimated threshold, specifically relating to the choice of econometric models and identification of causal effects. Although all the models passed the relevant tests of validity, there remained some uncertainty about the validity of the instruments. This structural uncertainty constituted a greater part of the overall uncertainty associated with the mortality effects of changes in expenditure, but the central estimate of the cost per QALY threshold was robust to this uncertainty.

The method of analysis used to link the effects of changes in expenditure on mortality to a fuller measure of health expressed in QALYs was also subject to uncertainty. A preferred analysis (or scenario) was identified as making the best use of available information, with assumptions appearing more reasonable than the available alternatives and providing a more complete picture of the likely health effects of a change in expenditure.

A critical issue is whether, on balance, the central or best estimate is likely to be an underestimate or overestimate of the cost per QALY threshold. Although other assumptions and judgements are possible that retain some level of plausibility, they do not necessarily favour a higher threshold. Indeed, when considered together, they suggest that, on balance, the central or best estimate of £12,936 is, if anything, likely to be an overestimate.

There are some reasons why the central estimate of the QALY threshold might be underestimated. First, in calculating life-year effects it is assumed that those deaths averted by a change in expenditure returns the individuals to the mortality risk of the general population (matched for age and gender). There are a number of other reasons why the central estimate might be overestimated. For example, the health effects of a change in expenditure are restricted to the population at risk during 1 year. This also means that the health effects of changes in expenditure which reduce incidence (prevention of disease) will not be captured either. A more formal and longer lag structure in the estimation of outcome elasticities would be likely to capture more health effects of a change in expenditure.

The effect of other assumptions that have been necessary are more ambiguous, although some evidence suggests their net effect may be conservative with respect to health effects of changes in expenditure.

The impact of investment, disinvestment and non-marginal effects

The central estimate of the cost per QALY threshold is based on estimates of the health effects of changes in expenditure across all 152 PCTs, some of which will be making investments (where expenditure is increasing) and others making disinvestments (where expenditure is reduced or growing more slowly).

The threshold is, however, likely to differ across these different types of PCT. It would be expected that, other things being equal, more expenditure would increase health but at a diminishing rate. Therefore, the amount of health displaced by disinvestment would be expected to be greater, and the associated threshold lower than the central estimate. Conversely, the health gained from investment would be expected to be lower, and the associated threshold higher.

This was examined by re-estimating the outcome and expenditure effects separately for those PCTs where their actual budget is under the target allocation from the Department of Health resource allocation formula (i.e. those under greater financial pressure and more likely to be disinvesting than investing), and those that are over target (under less financial pressure and more likely to be investing than disinvesting).

The results confirm these expectations. The health effects of changes in expenditure are greater when PCTs are under more financial pressure and are more likely to be disinvesting than investing. The analysis suggests that budget impact not only displaces more valuable activities within each PBC, but that overall expenditure tends to be reallocated to PBCs which can generate more health. Although further research might enable a quantitative assessment of how the relevant threshold should be adjusted for the scale of budget impacts, the qualitative assessment seems clear: the central estimate of the threshold is likely to be an overestimate for all technologies which impose net costs on the NHS (almost all technologies appraised by NICE); and the appropriate threshold to apply should be lower for technologies which have a greater impact on NHS costs.

How does the threshold change with overall expenditure?

The same methods were used to consider how the cost per QALY threshold is likely to have changed from 2007 to 2008 as overall expenditure has increased. This provides some insights into how the threshold might be expected to change over time as, for example, overall expenditure and NHS productivity changes.

This has implications for a judgement about the appropriate frequency of periodic reassessment of the cost per QALY threshold. Other things being equal, the threshold would be expected to increase following a rise in overall expenditure, although this will depend on whether or not there is discretion over how additional resources can be spent. However, insofar as the productivity of those activities that are valuable to the NHS also improves through innovation, the threshold will tend to fall. So, the net impact of these two countervailing effects on the threshold cannot be determined a priori.

Differences in the estimated thresholds between 2007 and 2008 were assessed. Although overall expenditure increased by 6% between 2007 and 2008, which represented real growth of 2% in 2007 prices, the overall threshold for all 23 PBCs fell by 5% in nominal terms and by 8% in real terms.

The reasons are complex but reflect changes in productivity, which differ across PBCs, but also a general reallocation of a change in overall expenditure towards those PBCs that appear more valuable in 2008. Given the uncertainty in estimation, subtle differences between 2007 and 2008 should not be overinterpreted. This analysis does suggest, however, that the overall threshold will not necessary increase with growth in the real or even nominal NHS budget. This suggests that the threshold is more likely to fall at a time when real budget growth is flat or falling and PCTs find themselves under increasing financial pressure.

What type of health is forgone by approval of a new technology?

The methods of analysis can identify not only how many QALYs are likely to be forgone across the NHS as a consequence of approving a technology which imposes additional costs on the NHS, but also where those QALYs are likely to be forgone and how they are made up (i.e. the additional deaths, life-years lost and the QoL impacts on those with disease).

As an example, based on the 2008 central estimate of the cost per QALY threshold (£12,936), the approval of ranibizumab (Lucentis®, Roche) for the treatment of diabetic macular oedema (prior to the Patient Access Scheme agreement) would have imposed additional annual costs of up to £80M on the NHS each year and been likely to displace 6184 QALYs elsewhere in the NHS. This forgone health is likely to be made up of 411 additional deaths and 1864 life-years forgone, most of which are likely to occur in circulatory, respiratory, gastrointestinal and cancer PBCs. However, much of the total health effect of these additional costs (4987 QALYs) is associated with QoL forgone during disease which is most likely to occur in respiratory, neurological, circulatory and mental health PBCs.

Conclusions and implications for practice

The research presented here goes some way to providing an empirically-based and explicit quantification of the scale of opportunity costs the NHS faces when considering whether or not the health benefits associated with new technologies are expected to be greater than the health that is likely to be forgone elsewhere in the NHS. As such, it provides a basis for determining the appropriate threshold for NICE decisions as well as those made centrally by the NHS and Department of Health more generally.

The methods presented can be used as a framework for further empirical work as additional and more appropriate data emerge in the NHS. They also offer a basis for threshold estimation in other health-care systems with budget constraints or limits on increasing expenditure.

The study also starts to make the other NHS patients, who ultimately bear the opportunity costs of such decisions, less abstract and more ‘known’ in social decisions. As who happens to be known or unknown is only a matter of perspective, time and ignorance, ethical and coherent social decisions require that both should be treated in the same way. These methods contribute to removing some of the ‘ignorance’ and making the unknown more real.

Research recommendations

There is a need to update estimates of the threshold with more recent and future waves of expenditure and mortality data.

If other aspects of social value are applied to health benefits of a new technology they must also be attached to the type of health that is likely to be forgone due to additional NHS costs. The methods developed here can be extended to allow weights to be also attached to the type of health that is forgone and estimate the wider social benefits that are likely to be lost when the NHS must accommodate the additional costs of new drugs.

We have demonstrated that these methods of analysis can be applied to QoL data collected as part of patient-reported outcome measures (PROMs). This type of analysis could be applied to these data in key PBCs as PROMs are rolled out providing some evidence about the QoL effects of changes in PBC expenditure.

A key PBC is mental health. Currently outcomes data that could be linked to measures of QoL are routinely collected in primary care. In principle, the same methods of analysis can be applied to these data once they are made available providing some evidence about the QoL effects of changes in mental health expenditure.

Improved and more recent estimates of incidence (by age and gender) and duration of disease will soon be available from the recently published updated World Health Organization Global Burden of Disease study. These data could be used when the threshold is re-estimated for later waves of expenditure data. Alternatively, estimates could be based on Clinical Practice Research Datalink data.

Estimating a more complex lag structure based on the evolving panel data would provide valuable evidence about the duration of the health effects of changes in expenditure. The recent release of census data for 2011 may allow a panel model to be estimated allowing better control for unobserved heterogeneity across PCTs as well as exploiting variation in outcomes, expenditure and other covariates over time. The formation of Clinical Commissioning Groups (CCGs) in 2013 will make the time series problematic for waves of expenditure after 2012 unless it is possible to match CCG and PCT boundaries.

If PBC expenditure and outcome data are available at CCG level (as well as covariates and suitable instruments), it might become possible to estimate outcome and expenditure equations simultaneously across PBCs. This would enable more of the likely health effects of changes in expenditure to be reflected in the analysis.

Funding

Funding for this study was provided by the NIHR-MRC Methodology Research Programme.

Health Technology Assessment

HTA MRP–MRC

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme or, commissioned/managed through the Methodology research programme (MRP), and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

This issue of the Health Technology Assessment journal series contains a project commissioned/managed by the Methodology research programme (MRP). The Medical Research Council (MRC) is working with NIHR to deliver the single joint health strategy and the MRP was launched in 2008 as part of the delivery model. MRC is lead funding partner for MRP and part of this programme is the joint MRC–NIHR funding panel 'The Methodology Research Programme Panel'.

To strengthen the evidence base for health research, the MRP oversees and implements the evolving strategy for high quality methodological research. In addition to the MRC and NIHR funding partners, the MRP takes into account the needs of other stakeholders including the devolved administrations, industry R&D, and regulatory/advisory agencies and other public bodies. The MRP funds investigator-led and needs-led research proposals from across the UK. In addition to the standard MRC and RCUK terms and conditions, projects commissioned/managed by the MRP are expected to provide a detailed report on the research findings and may publish the findings in the HTA journal, if supported by NIHR funds.

The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).