

**DOES NICE HAVE A COST-EFFECTIVENESS THRESHOLD AND WHAT
OTHER FACTORS INFLUENCE THEIR DECISIONS? A DISCRETE CHOICE
ANALYSIS**

NANCY DEVLIN
King's Fund, London

DAVID PARKIN
Department of Economics, City University, London

Paper presented to HESG, Brunel University, 3rd - 5th July, 2002

ABSTRACT

The decisions made by the National Institute of Clinical Excellence (NICE) give rise to two questions: how is cost-effectiveness evidence used to make judgements about the 'value for money' of health technologies? And how are factors other than cost-effectiveness taken into account? The aim of this study is to explore the cost-effectiveness threshold(s) and the tradeoffs between cost-effectiveness and other factors apparent in NICE decisions to date. Discrete choice analysis is used to reveal the preferences of NICE and to consider the consistency of its decisions. For each decision, the independent variables include: the availability (or not) of cost-effectiveness evidence; cost per life year gained; cost per Quality Adjusted Life Year gained; the net cost to the NHS; the burden of disease; the availability (or not) of alternative treatments; and specific factors indicated by NICE (for example, severity of condition, short duration of life). Data are drawn from the technical appraisal and guidance documents supporting the 39 decisions published by NICE from its inception to May 2002. This is work in progress: we welcome feedback on our model, hypotheses and initial results.

Introduction

The National Institute for Clinical Excellence (NICE) was established in 1999 to address variations in the use of pharmaceuticals between health economies in the UK (“postcode prescribing”) by providing national-level guidance on the effectiveness and cost-effectiveness of new health technologies in the NHS. To date it has issued guidance on 39 treatments; an ambitious programme of technical appraisals in support of further recommendations is planned for the future. The role of NICE has been strengthened by making implementation of its decisions mandatory in the NHS from 2002.

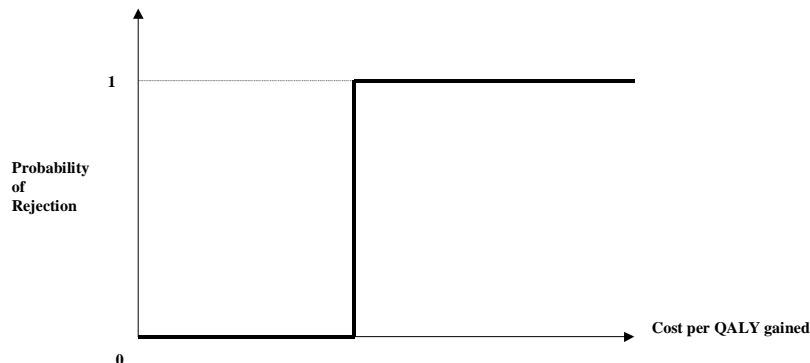
An issue generating considerable speculation and debate is the weight that NICE attaches to cost-effectiveness evidence in its decisions (Birch and Gafni, 2002, Raftery 2001) and, in particular, what decision rule it applies to incremental cost effectiveness ratios to judge whether any given treatment represents good value for money (Devlin and Towse, 2002). Towse (2002) has suggested that the ‘threshold’ cost per quality adjusted life year gained (CQG) *implicit* in NICE’s decisions is between £20,000 and £30,000; treatments with incremental cost-effectiveness ratios above this level seem more likely, but not certain, to be rejected. Explicit statements made by NICE are contradictory. The NICE guidance on Orlistat for obesity in adults (NICE 2002a) contained a statement that a “sufficient level of cost effectiveness” is “in the range of CQG of between £20,000 and £30,000”. Public comments made by the Chairman of NICE suggested that a threshold of £30,000 had emerged from its deliberations; however, NICE’s evidence to the Health Select Committee Inquiry maintains that these comments were misinterpreted and that “the Institute does not have such a threshold” (NICE, 2002b). Not only is there is no clear and explicit cost-effectiveness threshold, there is also a lack of clarity over the way in which factors *other* than health gain are taken into account - that is, the tradeoffs that are accepted between efficiency and objectives such as equity.

The aim of this paper is to consider the factors that operate to influence NICE decisions, to explore systematically the influence of each and to establish the characteristics of the cost-effectiveness threshold, if it exists. We report the results from initial data exploration and from a binary choice model using logistic regression analysis.

A binary choice model of NICE decision making

In terms of technical efficiency, an explicit threshold suggests a precise, 'knife-edge' value for a marginal QALY against which evidence from economic evaluation is compared: if the CQG exceeds this it is rejected; if it falls below it is accepted (Figure 1).

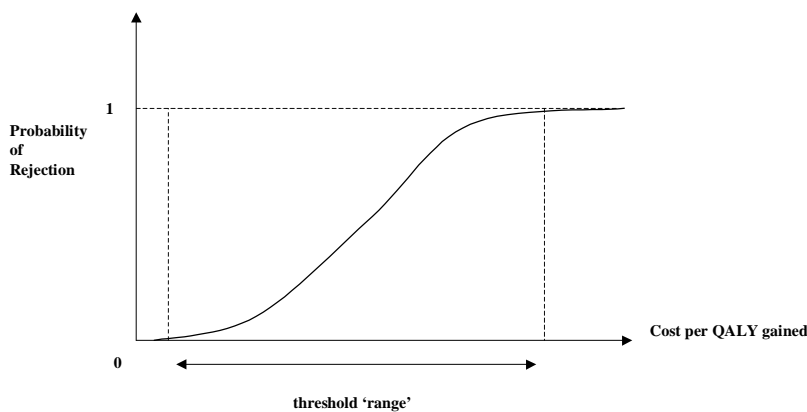
Figure 1
The cost-effectiveness threshold as a point



In practice, the threshold may be less clearly identified, for three reasons which we will consider in turn. First, decisions to accept or reject new technologies may depend on a wider set of objectives than maximising health gain from the NHS budget. Secondly, the cost effectiveness threshold may be different for investments and disinvestments. Thirdly, the threshold may be affected by the decision maker's response to uncertainty about evidence concerning cost-effectiveness.

The existence of factors other than cost-effectiveness may mean that there is in practice no threshold at all; any new technology has a finite probability of being accepted or rejected, whatever its CQG, if other factors are important enough to outweigh its cost-effectiveness. Alternatively, there may be no single threshold but instead a lower and an upper threshold (Figure 2). Below the lower threshold, low CQG technologies are certain to be accepted; above the upper threshold, high CQG technologies are certain to be rejected. *Within* the range between the two, cost effectiveness may be traded off against other objectives that are seen as relevant to decision making.

Figure 2
The cost-effectiveness threshold as a range, reflecting tradeoffs against efficiency



What are the factors other than cost effectiveness that may influence NICE decisions? An obvious candidate is equity (Hutton and Maynard, 2000; Birch and Gafni, 2002); less obvious is *what* equity concerns are relevant to NICE’s deliberations. The NHS has as one of its central equity principles access to health care irrespective of ability to pay. This is addressed by the means by which treatments recommended by NICE are funded - general taxation, free at the point of delivery - and is therefore not relevant to its deliberations. Equal access regardless of geography is also an important equity consideration - the avoidance of the ‘postcode lottery’ (regional variation in access to some treatments) was, as already noted, a

principal objective in establishing NICE. This is mainly addressed by weighted population based funding formulae, which aim to ensure that local health organisations have equal resources for equal need. NICE's role is in effect to ensure that this equal availability of resources is translated into equal availability of specific therapies. However, it is again hard to see how it could affect NICE's decisions, as its recommendations pertain to the NHS as a whole. Population characteristics *other* than income and geography, for example age, sex, ethnicity and social class, are possible foci of equity goals. But we do not believe that these are relevant equity criteria for NICE, as it has no mandate for differentiating between population sub-groups (e.g., by weighting QALYs¹) and, arguably, discrimination legislation precludes its ability to do so on some grounds.

While the relevance of equity goals regarding *population* sub-groups can be largely rejected *a priori*, by contrast, concerns about equity between *patient* groups is likely to be highly relevant to NICE's decision making processes. We consider three ways in which this might be implemented. First, NICE might approach 'orphan' treatments (i.e., a treatment for a disease for which no alternative curative treatment for patients exists) differently from treatments for which there *are* treatment alternatives. Secondly, the 'starting point' in health status terms of particular patient groups - low initial quality of life and short duration of life - may be seen as a relevant and inadequately captured by the measures of health *gain* used in economic evaluation. Thirdly, cost effectiveness *ratios* do not differentiate between the size of the potential group of beneficiaries. A larger patient group, where the total health gain produced from a treatment is larger, may be approached differently from a smaller patient group. In each case, treatments may have a lower probability of being rejected for any given CQG.

A final other possible factor is related neither to efficiency nor equity, but is suggested by the evidence that NICE requires for its technical appraisals on the net

¹ However, the choice of QALYs (and LYG) as the measure of benefit in cost effectiveness analysis in itself embodies value judgements about the value of health gains between age groups.

budgetary effect of its guidance on the NHS. The role that this evidence does or should have in decision-making is an area of dispute. In one view, NICE makes its decisions on the basis of effectiveness and cost effectiveness evidence, with information on the impact on the NHS budget being used only to operationalise those decisions; that is, to plan how much the NHS budget would need to increase in total to support a new procedure, or what resources will need to be displaced at a local level to implement guidance. Birch and Gafni (2002) argue that "...the puzzle here is how recommendations can be made for maximising health gain from a given NHS expenditure where such recommendations require additional resource requirements (and of unknown opportunity cost). If budgetary impact is only important in planning future resource requirements then all interventions with net benefits² would be implemented and NICE recommendations would be a prescription for continued expansion of the NHS" (p.189). In practice, the establishment of NICE has coincided with a period of unprecedented, planned increases in real NHS budgets, making it impossible to determine cause and effect from casual observation alone. However, as Raftery (2001) notes, NICE has said 'yes' more often than it has said 'no'.

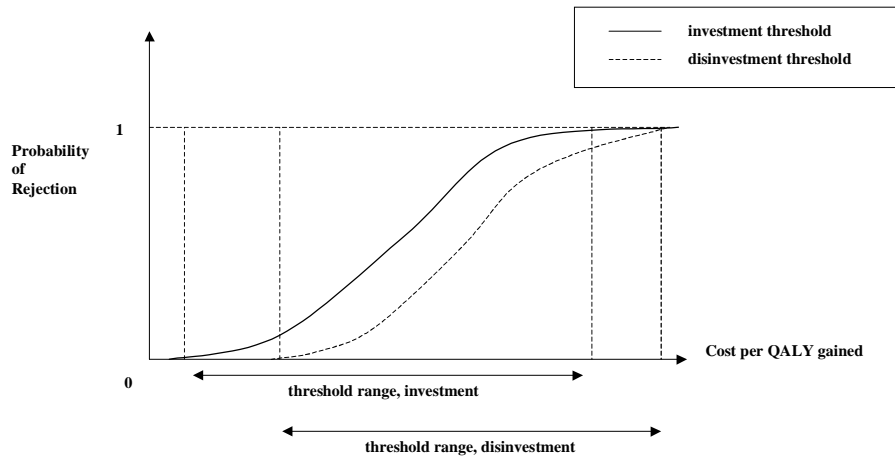
The second complicating factor in looking at a possible cost effectiveness threshold is that it may be different for investments and disinvestments. O'Brien *et al* (2002) provide evidence that the willingness to accept (WTA) values for relinquishing QALYs, by reducing or removing services, are higher than the willingness to pay (WTP) to obtain QALYs from "new" services³. This suggests that the cost

² This begs the question of what 'net benefit' means in this context. NICE's use of cost effectiveness, rather than cost benefit, analysis means that net benefit can only be determined by invoking a threshold. If the threshold reflects the existing NHS budget (e.g., if it is revealed by the CQG of the last treatment funded from the NHS budget - the 'extra welfarist' position) rather than an externally-determined, stated value per QALY gained (the 'welfarist' approach - see O'Brien et al 2002) this argument becomes invalid.

³ The asymmetry described by O'Brien et al (2002) is arguably even more exaggerated in NICE decisions since its remit is to consider cost effectiveness *and* clinical effectiveness (Department of Health 1999). While NICE recommendations in favour of new technologies indicates its WTP for increased effectiveness, it is less clear how it would respond to an option which is cost effective because it is both less costly *and less effective* than current practice. If cost-effectiveness and effectiveness are equally dominant in NICE's preference function, its WTA would be infinitely large i.e., there would be no reduction in costs sufficient to compensate for reduced effectiveness. We do not test for this possibility here, as all options considered by NICE to date sit in the North-East and South East quadrants of the cost effectiveness plane described by O'Brien et al (2002) i.e. have zero to positive effectiveness.

effectiveness threshold may be lower at every given CQG for extant, as opposed to new, services (Figure 3).

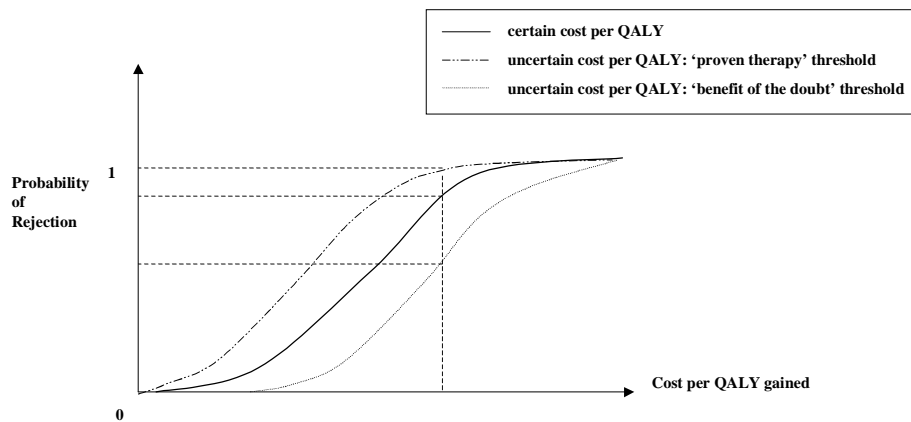
Figure 3
The cost-effectiveness threshold for investments and disinvestments



The final complicating factor is that the decision maker's response to uncertainty regarding CQG evidence arising, for example, from sensitivity analysis, may alter the threshold. If NICE is risk averse, the probability of rejection will be higher for any given base-case CQG for options associated with the possibility of a high CQG under alternative sets of assumptions, compared to options where the base case CQG is relatively robust to changes in assumptions. If NICE is a risk lover, it will be prepared to give "the benefit of the doubt" and the opposite will apply (Figure 4).

Figure 4

The cost-effectiveness threshold under uncertainty



In terms of a binary choice model, the response variable is the probability that NICE will reject a given therapy. In some cases, the NICE Guidance document involves a simple acceptance or rejection of a treatment for “routine use”, for example recommending that a drug should be available to all sufferers of a particular condition. However, many NICE Guidance documents specify both clinical groups for whom the treatment is recommended and those for whom it is not (Raftery, 2001). In these cases, the guidance actually implies more than one decision: acceptance for some identified groups and rejection for others.

Our exposition of the threshold in the previous section immediately suggests a number of explanatory variables. Incremental cost effectiveness ratios are cited in many of the Guidance documents and in the underlying Technical Appraisal documents. The relevant equity variables can be measured by examining the characteristics of patient subgroups. The hypothesis that NICE is in fact indifferent to the cost impact of its guidance can be tested by including information on the budgetary impact on the NHS. The effect of uncertainty over the cost-effectiveness evidence may be measured by exploring the range of cost effectiveness results reported in the Guidance supporting each decision.

Data

Data were abstracted from NICE Guidance and Technology Appraisals (TA) available at May 2002, covering recommendations on 39 therapies, corresponding to 51 observable yes/no decisions. The abstracted data are presented in Appendix 1. In most cases these data are self-explanatory - however, note that the impact on the NHS is that of *approving* the treatment and is not identical to the cost of the actual NICE decision, which may either have approved *or* rejected the treatment.

The quantitative variables used in modelling are summarised in Table 1.

Table 1: Summary of variables

<i>Variable name</i>	
<i>Dependent variable:</i>	
DECISION	A binary choice variable which takes the value 0 if the decision is in favour of use, 1 if against.
<i>Independent variables:</i>	
CER	Cost per quality adjusted life year gained or cost per life year gained in £1,000 units. Where a single estimate or a central or base case estimate is provided, this is used. If only a range is given, the mean is used.
UNCERTAINTY	Uncertainty about the cost-effectiveness evidence, calculated as the range of cost effectiveness ratio divided by the mean or base case cost-effectiveness ratio.
OTHER THERAPY	A dummy variable set to 1 if there are no treatment options and 0 if alternative treatments are available as a substitute for the treatment under consideration.
OTHER FACTORS	A dummy variable set to 1 if the guidance report makes specific mention of other variables influencing its

	recommendation (severity of condition, short duration of life etc), 0 otherwise.
BURDEN	The burden of disease, represented by the number of people affected by the condition to which the option pertains, in 1,000 units.
IMPACT	The impact on NHS budgets of approving the treatment in £1m units. Where the net impact is available this is used, otherwise it is the impact of the therapy itself. Where a range is given, the mean is taken.

A number of issues must be discussed about the data used for modelling. The cost-effectiveness ratio (CER) was the cost per QALY gained (CQG) where it is reported and the cost per Life Year gained (CLYG) where it is not. (The guidance for some decisions reports both.) This implies a one-to-one correspondence between life years gained (LYG) and Quality Adjusted Life Years gained (QALYG). This equivalence will be incorrect if those therapies for which CLYG is reported either have an additional impact, positive or negative, on existing quality of life or produce life years of less than full quality, or both.

Although the main reason for doing this is to produce sufficient data for analysis, there are some justifications. First, in most cases the guidance for therapies for which only CLYG is reported either explicitly say that there are no quality adjustments to be made or implicitly do so by not mentioning this factor. This implies that NICE believes that the evidence is that LYG and QALYG are in fact the same in those cases. Secondly, an assumption of this analysis is that NICE decision makers respond mainly to data that they are presented with. In the absence of any data on quality adjustments to LYG, which could of course raise or lower the CQG, the lack of such quality adjustments may not be noticed or taken into account. Thirdly, a slighter weaker justification is that we are especially interested in the decisions that are “out-of-order” with respect to CERs, and these apparent anomalies are the same in the two subsets as they are in the data as a whole.

Another problem with the CERs is that they use inconsistent perspectives in costing. Some include patient costs (for example the technical appraisal of beta interferon for multiple sclerosis), some use public sector costs other than those to the NHS (for example, social care costs in the technical appraisal for Trastuzumab), still others restrict their perspective to the NHS⁴. This suggests that the CQG cannot directly be compared between them and, indeed, that a different threshold would apply in each case. Again, however, these were the data which were considered by NICE and there is no evidence that the appraisal committees adjusted these formally or informally to convert them to a common bases for their decision making.

The variable UNCERTAINTY would be better represented by the coefficient of variation of the CER. However, the data are rarely if ever presented in the form of a distribution from which a true mean and standard deviation can be calculated. Hence we had to use the range as a measure of spread and the base case or average of the range as a measure of central tendency.

Appendix 1 does not present the data for the equity variable OTHER FACTORS: in only three cases – the treatments considered for Motor Neurone Disease (MND), pancreatic cancer and non-small cell lung cancers - did a Guidance make particular reference to health status ‘starting point’ issues in its decision. The clearest indication of factors other than cost-effectiveness influencing its deliberations is provided by the guidance for Riluzole for MND, in which it is noted that “The Committee took account of the severity and relatively short life span of people with ALS and, in particular, as directly reported to it, of the values which patients place on the extension of tracheostomy free survival time” (NICE, 2001b, p.5). The Guidance for treatments for pancreatic cancer and non-small lung cell cancer refer to the “extremely poor prognosis” and low survival rates in each case (although CQG evidence is not available in either decision).

⁴ NICE’s guidance to manufacturers and sponsors for preparing evidence (NICE 2001a) now provides clear advice on the perspective to be taken in economic evaluation.

The variable BURDEN covers four different, and incompatible, measures of the number affected. These are the number of cases, new cases, treatments and deaths. Again, we have combined these mainly in order to provide sufficient cases for analysis, but again there is a weak argument that NICE decision makers may respond to the magnitudes presented and may not distinguish too finely between the different types of burden.

The variable IMPACT, representing the budgetary implications for the NHS also suffers from inconsistent evidence provided to and by NICE. Some are not incremental (for example, drug costs are 'totals', not taking into account existing spending on that drug where it is already used in the NHS); others take partial account of changes in resource use (for example, where use of one drug or procedure replaces another); and others are based on a fuller account of changes in resource use (for example, changes in hospital use arising from longer term changes in morbidity). We have had to treat these comprehensive estimates of net impact as equivalent to estimates of the direct cost of the therapy and to make estimates of the average where only ranges are given. Again, there is only a supporting argument that magnitudes are what are visible and may enable differences in meaning to be neglected.

A more general point is that abstraction of the data is difficult not only because of the complexity of the evidence in the guidance and technical appraisal reports and the many differences between the data presented in them, but also because it is not clear which data were actually taken into account by NICE. There is in many cases a disparity of some magnitude between the CQG reported in the Guidance and that reported in the supporting technical appraisal. Previous analyses of the role of cost effectiveness evidence in NICE decisions have restricted their analysis to the evidence reported in the Guidance (Towse, 2002; Raftery, 2001). While it is to be expected that the Guidance committee would consider factors other than cost effectiveness in issuing their recommendations, the committee also supplements the

independent technical appraisals with confidential manufacturers' evidence as well as incorporating more 'casual' economic evidence and reasoning. In some cases, the CQG reported in the Guidance is the committees' "best guess" about cost effectiveness, taking into account factors they consider not to be included in the technical appraisal proper.

For example, the Guidance for Riluzole for MND cites a CQG for this treatment of £34,000-£43,500. The technical appraisal has a base case CQG of £58,000, with sensitivity analysis revealing a range of CQG from negative to a considerably higher than base-case CQG. Later revisions to the technical appraisal, in the light of new evidence, suggest the CQG ranges from £16,500-£20,000. The wide range of results is illustrative of the degree of uncertainty NICE faces in using cost-effectiveness evidence. However, for the purpose of this analysis, the key point to note is that the CQG figures cited in the Guidance cannot be located in the evidence in any of the technical appraisals it has provided in support of its decision.

A second example is the cost-effectiveness data for Orlistat for obesity in adults. As stated earlier, the NICE guidance for Orlistat seemed to admit to an approximate threshold of £20,000-£30,000 per QALY gained and it further implies, by its favourable decision, that Orlistat meets this. However, the "headline" estimates that it gave for this CER are a much higher independent estimate of £46,000, which is neither endorsed nor rejected, and an explicitly rejected much lower figure based on manufacturers estimates.

These "headline" figures are taken from the technical appraisal, which itself makes them the "headline". But the technical appraisal also reports, in a less prominent way, some sensitivity analyses around the figures, which are not referred to in the guidance and which support a lower CER. There is however, nothing within those analyses which supports the figure of £20,000-£30,000.

The technical appraisal takes the independent estimate from a Development and Evaluation Committee (DEC) report (The Wessex Institute 1999), which again headlined the £46,000 figure, but also contains far more sensitivity analyses than those reported in the technical appraisal. These also support a far lower CER. Of particular note is that the headline figure is based on the DEC method of calculating QALY gains, but the DEC report also gives a much lower set of CER estimates based on EQ-5D utilities, explicitly noting that NICE appears to be preferring estimates based on the EQ-5D. These CERs did not find their way into the technical appraisal and it is unclear whether or not they will have been considered by NICE. However, once again, there is no obvious source in the DEC report for the £20,000-£30,000 figure.

The conclusion is that there is some uncertainty in many cases about what NICE's conclusions about cost-effectiveness were, the means by which they were derived and indeed what evidence they took into account in deriving them. Nevertheless, we are obliged to take them at face value.

Methods

The decisions were initially divided into those for which cost-effectiveness data was available and those for which there was none. The latter were subdivided into acceptance and rejection decisions and these were investigated qualitatively to uncover the apparent reasons for rejection or acceptance. The former were amenable to quantitative analysis of a binary choice model, which was explored using logistic regression analysis.

Several model specifications incorporating different numbers of variables were estimated. However, this was for reasons other than a model building strategy. First, our intended strategy was to estimate a model incorporating all variables, but this could not be done, for reasons explained below, and alternative specifications including fewer variables had to be examined. Secondly, models incorporating

fewer variables were in themselves instructive about the impact of the implied decision making criteria. Because we are not estimating a full model, robust standard errors were used in all specifications as a basis for hypothesis testing.

The modelling was constrained by the absence of some items of data. The variable OTHER FACTORS could not be used, because the decision for all but one of the cases for which other factors were recorded was to accept and, for the other, data on cost-effectiveness were not available; OTHER FACTORS therefore had no explanatory power. One decision (27a) was excluded because of missing burden of disease data, but this was not an unusual case and had little effect on the results. The variable IMPACT was also available only for a restricted set of decisions, which unfortunately included four of the seven rejections; moreover these were amongst the more interesting cases. As a result, we used a basic set of 33 decisions, although we also used a subset of 26 to test the IMPACT variable.

Results

Table 2 details the decisions for which there were no cost-effectiveness data, divided into those accepted and those rejected. Although the guidance is in many cases complex, the general conclusion, as might be expected, is that those rejected are those for which there is clear evidence that the therapy is *not* effective, for example decision 1, or very unclear evidence that it *is* effective, for example decision 23. The technologies that were accepted without evidence on CLYG or CQG fell into two groups. In five cases, the decision is arguably better characterised as *which* treatment of those considered is most appropriate (for example, which type of prostheses should be used in hip replacement, given their differences in cost and revision rate), rather than whether any treatment *per se* is acceptable value for money relative to other NHS activities. Decisions reflect cost-minimisation or effectiveness-maximisation. In the remaining four cases, cost effectiveness evidence was considered, but in terms other than CLYG or CQG (for example, cost per year of remission).

Table 3 shows separately those decisions for which there was evidence in terms of CLYG and of CQG. These are ordered from lowest to highest CER, with rejection decisions shown in bold. As may be seen, decisions based on CQG were broadly in the expected order except for decision 15, supporting a “lower threshold” of between £37,500 and £38,000, and an “upper threshold” between £46,000 and £50,000. Decisions based on CLYG appear to be much less well-ordered.

An attempt was made to estimate a logistic regression model including all of the usable variables except OTHER FACTORS, for reasons explained earlier. Unfortunately, this proved not to be estimatable, providing completely determined outcomes along with odds ratios and standard errors which had bizarre signs and magnitudes.

Table 4 shows the results which were obtained from the logistic regression analyses; Table 5 shows how these affect both the implied NICE “threshold” and which decisions conform to it. The first section of Table 5 is a simple listing in ascending order of cost-effectiveness ratios (CERs), with rejection decisions shown in bold. Those with the three highest CERs (18, 27 and 32) are rejections. Rejection decision 5 is a clear outlier. However, it is less clear whether it is rejection decisions 30, 33 and 15 or the fourteen acceptance decisions within and above them that are “out of order”.

Model 1 in Table 4 adds UNCERTAINTY to CER as an explanatory variable. Both have odds ratios which are significant and of the expected signs. The impact of this on the ranking of decisions is shown in the second set of columns of Table 5, where decisions are sorted by the calculated probability of rejection. Three observations may be made. First, rejection decision 5 is no longer an outlier, having the third highest rejection probability. The UNCERTAINTY variable appears to explain well this rejection despite a low CER, which is consistent with the reasons for this rejection stated by the NICE guidance. Secondly, it is now possible to observe a

“threshold” as the point at which the probability of rejection becomes greater than the probability of acceptance, other things being equal. This lies between £38,750 and £50,000. Thirdly, it is also now possible to classify which decisions should be viewed as “out of order”. These are rejection decisions 15, 30 and 33, because they have rejection probabilities below 0.5, and acceptance decision 22 because it has a rejection probability greater than 0.5.

Decision 32 is reported as having a probability of 1, although strictly speaking it is less than one but by a smaller amount than the computer can handle. This decision is a clear outlier because of its very high estimated CER and all models have the same result. An important point about this is that it has no detectable influence on the estimated odds ratios and only an extremely small influence on the standard errors in all of the models. It therefore also has no detectable influence on the estimated probabilities of other decisions.

Although the UNCERTAINTY variable seems to explain rejection decision 5 and is consistent with statements within the guidance, closer examination shows a less clear picture. The UNCERTAINTY variable measures the degree of variability about the CER, which is certainly relatively high in this case. However, the absolute value of the high estimate of the CER is still well below the implied threshold. The implication is either that NICE’s decision took account of the presence of high uncertainty but ignored the significance of this uncertainty, or that the level of uncertainty actually attached to the estimate by NICE was far higher than that reported. We have no way of judging which of these is closer to the truth.

Model 2 in Table 4 adds the variable BURDEN to Model 1 to assess its incremental impact. An alternative is also to examine a model including CER and BURDEN alone, but the estimated odds ratios for that model were not significant. In Model 2, all are significant and have the expected signs; those for CER and UNCERTAINTY are slightly higher than in Model 1 but have similar standard errors. The impact is seen in the third set of columns in Table 5. The most important change from the

Model 2 table is that no acceptance decisions are “out of order”; acceptance decision 22 now has a probability of rejection below 0.5, because of the very large number of people affected. The “out of order” rejection decisions remain and the implied threshold is the same.

Model 3 adds the variable OTHER THERAPY. Table 4 shows that the odds ratios are again all significant and have the expected signs. CER, UNCERTAINTY and BURDEN have similar odds ratios to those in Model 2, but the standard errors are slightly higher. The fourth column of Table 5 shows that this model has some important effects. The implied “threshold” is now rather lower, lying between £25,000 and £37,500. Rejection decision 15 is now no longer “out of order”; however rejection decisions 30 and 33 remain so. In addition, there are now two “out of order” acceptance decisions, numbers 22 and 34.

Model 3 needs to be interpreted carefully. The estimated odds ratio for OTHER THERAPY is extremely large and probably not meaningful. Its effects can be observed in Table 5, where the five decisions affected by this variable are shown in italics. Four are given rejection probabilities approaching zero and the other decision is number 32 for which the probability, as always, approaches one. However, the same issue arises as with the inclusion of number 32; the inclusion of these five decisions has no influence on the odds ratios of variables other than OTHER THERAPY or on the estimated probabilities of other decisions. The conclusion concerning the threshold is therefore robust.

The χ^2 statistic suggests that none of these three models can be rejected. Model 3 is preferred because of its higher Pseudo R² and better sensitivity, achieved at the cost of a slightly lower specificity⁵.

⁵ The same odds ratio estimates are obtained if the decisions for which there is no other therapy and the other therapy variable are excluded, but of course the goodness of fit statistics are different. However, the same conclusions about them hold.

Model 4, shown in Table 4, adds the variable IMPACT to Model 2. This is a less useful model, but is shown for illustration. As explained, it requires a more restricted data set that includes only three rejections, all of which are fully explained by the CER and UNCERTAINTY variables. There is a large effect on the odds ratio of UNCERTAINTY, both BURDEN and IMPACT have an unexpected sign, though neither is significant, and the model is completely determined, with no “out of order” decisions. Its impact is therefore not shown in Table 5. A further model was estimated adding both OTHER THERAPY and IMPACT to Model 2, but this was also not properly estimatable, again producing a completely determined model with odd odds ratios and erroneous standard errors.

Discussion

The following tentative conclusions can be drawn from the analysis of NICE decisions for which there were cost-effectiveness data. There is support for the idea of a threshold as being probability based rather than a single number. NICE decisions are well explained by the cost-effectiveness evidence, with the effect of uncertainty and of the burden of disease explaining the rejection of some therapies with a relatively low CER and the acceptance of some with a relatively high CER. Estimates of the threshold itself are additionally affected by the fact that some of the conditions for which therapies have been accepted with a relatively high CER have no alternative therapies. There remain a few anomalies, which may be the result of the context in which these decisions were taken; essentially that they were decisions based on comparisons of different indications for particular therapies rather than taking into account a comparison with other services that the NHS provides. The analysis supports a cost-effectiveness threshold consistent with the £20,000-£30,000 figure which NICE has publicly identified but denied.

The modelling results are of course preliminary and exploratory and are far from definitive. A key problem is of course the data used, which are less than perfect. However, the decision making that we are trying to model presumably had the same

imperfect data. Unless NICE decisions are based upon data which are not made available to the public, which is definitely the case for manufacturers' commercial-in-confidence data, then the modelling is using the correct data, though of course they may not have been interpreted by the NICE appraisal committees in exactly the same way as we have done.

The analysis may have revealed as much about the data upon which decisions have been based as about the decision making itself. NICE should not be criticised for making decisions based on imperfect and missing data; its role is to exercise judgement where that is the case. Of more concern are the widespread inconsistencies in the type of evidence collected and reported, the mismatch between figures reported in the guidance, the technical appraisals and other documents and the obvious existence of key documents and analysis which are not in the public domain. A consequence of this is a reduction in the explicitness and transparency of the decision-making process.

The insights from this work will provide a systematic way of identifying the role of various types of evidence on decision making retrospectively. Inferring the threshold from past decisions usefully serves to promote debate about the threshold and about how NICE should respond to equity concerns. However, the question of what NICE's cost-effectiveness threshold *should* be, as opposed to what it appears to have *been*, is more challenging still and ultimately needs to be supported by stated, rather than implicit, valuations of health outcomes.

References

Birch, S., Gafni, A. (2002) On being NICE in the UK: guidelines for technology appraisal for the NHS in England and Wales. *Health Economics* 11:185-191.

Department of Health (1999) *Faster access to modern treatment: how NICE will work*. London: HMSO.

Devlin, N., Towse, A.(eds) (2002) *Cost effectiveness thresholds: economic and ethical issues*. London: King's Fund/Office for Health Economics.

Hutton, J., Maynard, A. (2000) A NICE challenge for health economics. Editorial. *Health Economics* 9:89-93.

NICE (2002a) *Guidance on the use of Orlistat for the treatment of obesity in adults*. Technology Appraisal Guidance No. 22.
www.nice.org.uk

NICE (2002b) *Health Select Committee Inquiry: submission of supplementary evidence. Part 2 of 2*.
www.nice.org.uk/Docref.asp?d=27652

NICE (2001a) *Guidance for manufacturers and sponsors*. Technology Appraisals Process Series 5. June 2001.

NICE (2001b) *Guidance on the use of Riluzole for the treatment of Motor Neurone Disease*. Technology Appraisal Guidance No. 20.
www.nice.org.uk

O'Brien, B.J., Gersten, K., Willan, A.R., Faulkner, L.A. (2002) Is there a kink in consumers' threshold value for cost effectiveness in health care? *Health Economics Letters* 6 (1).

Raftery, J. (2001) NICE: faster access to modern treatments? Analysis of Guidance on health technologies. *British Medical Journal* 323:1300-1303.

The Wessex Institute (1999) *Orlistat for the treatment of obesity*. Development and Evaluation Committee report No. 101.

Towse, A. What is NICE's threshold? An external view. Chapter 2 in: Devlin, N., Towse, A.(eds) (2002) *Cost effectiveness thresholds: economic and ethical issues*. London: King's Fund/Office for Health Economics.

Table 2: NICE decisions for which there were no cost-effectiveness data

ACCEPTED		REJECTED	
Number	Therapy	Number	Therapy
2	Hip	1	Wisdom
7	Proton pump	16	Cartilage
8	Hearing aid	17	Laparoscope colorectal
9	Rosiglitazone	23	Temozolamide (First line)
10	Inhalers	25	Gemcitabine (Other)
21	Pioglitazone	28	Topecetan No
24	Wound care	33	Advanced colorectal 4
29	Fludarabine	37	Rituximab
33	Advanced colorectal 2		

Table 3: NICE decisions ranked by cost-effectiveness ratio. (Rejections in bold.)

Number	Therapy	CER
39	Smoking	430
28	Topecetan Yes	1000
5	Cytology	1100
38	Asthma inhalers	5000
3	Taxane Ovarian	8271
12	Glycoprotein	9250
26	Non-small cell lung (First line)	9475
27	Cox II High risk)	11244
13	Methylphenidate	12500
25	Gemcitabine (First line)	12950
26	Non-small cell lung (other)	14000
19	Alzheimers	15000
30	Taxane Breast 2 (Second line)	15250
6	Taxane Breast	15500
30	Taxane Breast 2 (First line)	19000
34	Trastuzumub (monotherapy)	19000
15	Zanamavir At Risk	20400
14	Ribavarin	20500
33	Advanced colorectal 3	22500
31	Sibutramine	22500
35	Arthritis juvenile	22500
18	Laparoscope hernia (recurrent)	25000
4	Stents	25000
11	ICDs	28500
33	Advanced colorectal 1	29000
36	Arthritis adult	31000
23	Temozolamide (Second line)	35000
34	Trastuzumub (combination)	37500
15	Zanamavir All	38000
20	Riluzole	38750
22	Orlistat	46000
18	Laparoscope hernia (primary)	50000
27	Cox II (Routine)	150000
32	Multiple sclerosis	307500

Table 4: Logistic regression analyses of NICE decisions

	Model 1		Model 2		Model 3		Model 4	
CER	1.108109	(.0558801)	1.150873	(.057253)	1.192245	(.075416)	1.253999	(.075608)
UNCERTAINTY	2.162434	(.6292686)	2.538829	(.728284)	2.693212	(.806602)	616.798	(167.274)
BURDEN			.9927932	(.003269)	.9899056	(.004367)	<i>1.012854</i>	<i>(.007830)</i>
OTHER THERAPY IMPACT					3.00e+08	(5.80e+08)	<i>.9552886</i>	<i>(.065333)</i>
N	33		33		33		26	
Log likelihood	-10.544938		-9.518633		-8.108971			
Pseudo R ²	0.3816		0.4418		0.5245			
Pearson χ^2	26.66	(0.5901)	23.33	(0.7611)	18.65	(0.9086)		
Sensitivity	57.14%		57.14%		71.43%			
Specificity	96.15%		100.00%		92.31%			
Correctly classified	87.88%		90.91%		87.88%			

Robust standard errors for odds ratios are shown in parentheses.

Probability values of the Pearson χ^2 Goodness of fit statistics are shown in parentheses.

Table 5. Ranking of NICE Guidance decisions by cost-effectiveness and probability of rejection.

CER		Model 1			Model 2			Model 3		
	CER		CER	Prob		CER	Prob		CER	Prob
39	430	28	1000	.00763	38	5000	.00002	<i>14</i>	<i>20500</i>	<i>.00000</i>
28	1000	38	5000	.01147	39	430	.00312	<i>19</i>	<i>15000</i>	<i>.00000</i>
5	1100	39	430	.01318	28	1000	.00369	<i>23</i>	<i>35000</i>	<i>.00000</i>
38	5000	12	9250	.02546	12	9250	.00838	<i>20</i>	<i>38750</i>	<i>.00000</i>
3	8271	3	8271	.02761	36	31000	.01575	38	5000	.00000
12	9250	26	14000	.02840	13	12500	.01639	39	430	.00138
26	9475	13	12500	.03299	26	14000	.01857	28	1000	.00222
13	12500	34	19000	.04657	3	8271	.01999	12	9250	.00506
25	12950	25	12950	.04949	14	20500	.02727	36	31000	.00836
26	14000	26	9475	.05622	19	15000	.02845	13	12500	.01249
19	15000	30	15250	.07112	26	9475	.03984	3	8271	.01631
30	15250	6	15500	.07361	25	12950	.04360	26	14000	.01634
6	15500	30	19000	.08552	30	19000	.08782	26	9475	.03286
30	19000	33	22500	.09257	6	15500	.06089	25	12950	.04219
34	19000	31	22500	.10475	18	25000	.25411	34	19000	.05015
15	20400	35	22500	.10475	30	15250	.06636	6	15500	.06017
14	20500	4	25000	.10962	34	19000	.04448	4	25000	.06621
33	22500	15	20400	.11547	31	22500	.09647	30	15250	.06851
31	22500	33	29000	.12000	11	28500	.09546	30	19000	.10083
35	22500	11	28500	.12915	33	22500	.09232	33	22500	.11168
18	25000	19	15000	.13157	35	22500	.12097	31	22500	.11285
4	25000	14	20500	.15463	15	20400	.13723	11	28500	.11399
11	28500	36	31000	.16965	33	29000	.13852	35	22500	.15371
33	29000	23	35000	.20157	4	25000	.06440	15	20400	.16918
36	31000	34	37500	.24604	23	35000	.31506	33	29000	.19543
23	35000	15	38000	.25568	34	37500	.38518	18	25000	.28719
34	37500	18	25000	.29721	15	38000	.39070	34	37500	.57734
15	38000	20	38750	.30950	22	46000	.47705	15	38000	.58270
20	38750	18	50000	.54074	20	38750	.49135	22	46000	.60125
22	46000	22	46000	.58814	18	50000	.63942	18	50000	.81838
18	50000	5	1100	.79955	5	1100	.87841	5	1100	.87629
27	150000	27	150000	.99997	27	150000	.93859	27	150000	.91679
32	307500	32	307500	1	32	307500	1	32	307500	1