

# Papers

## Comparing estimates of cost effectiveness submitted to the National Institute for Clinical Excellence (NICE) by different organisations: retrospective study

A H Miners, Martina Garau, Dogan Fidan, A J Fischer

### Abstract

**Objective** To assess the association between different types of organisation and the results from economic evaluations.

**Design** Retrospective pairwise comparison of evidence submitted to the technology appraisal programme of the National Institute for Clinical Excellence (NICE) by manufacturers of the relevant healthcare technologies and by contracted university based assessment groups.

**Data sources** Data from the first 62 appraisals.

**Main outcome measure** Incremental cost effectiveness ratios.

**Results** Data from 27 of the 62 appraisals could be compared. The analysis of 54 pairwise comparisons showed that manufacturers' estimates of incremental cost effectiveness ratios were lower (suggesting a more cost effective use of resources) than those produced by the assessment groups (25 were lower, 29 were the same, none were higher,  $P < 0.01$ ). Restriction of this dataset to include only one pairwise comparison per appraisal (27 pairs) produced a similar result (21 were lower, two were the same, four were higher,  $P < 0.001$ ).

**Conclusions** The estimated incremental cost effectiveness ratios submitted by manufacturers were on average significantly lower than those submitted by the assessment groups. These results show that an important role of NICE's appraisal committee, and of decision makers in general, is to determine which economic evaluations, or parts of evaluations, should be given more credence.

### Introduction

Evidence suggests that profit making organisations are more likely to report favourable results and conclusions from clinical studies than non-profit organisations. The most recent of these studies concluded that for profit organisations were 5.3 (95% confidence interval 2.0 to 14.4) times more likely to recommend the use of an experimental drug compared with non-profit groups.<sup>1</sup> Only three studies, however, have formally assessed this association with regard to the results from economic evaluations. One showed that evaluations of oncology products sponsored by drug companies were less likely to report unfavourable qualitative conclusions (5% *v* 38%;  $P = 0.04$ ) compared with studies sponsored by non-profit organisations.<sup>2</sup> The two other studies also reported similar findings.<sup>3-4</sup>

The technology appraisals programme of the National Institute for Clinical Excellence (NICE) provides guidance to the NHS in England and Wales on the use of new and existing health technologies. Manufacturers (or relevant UK agent) of the

relevant technology, professional groups, and national groups representing patients submit evidence. Collectively, these three groups are known as consultees. An academic centre commissioned through the NHS's health technology assessment programme, called an assessment group, also assesses the evidence. Guidance to the NHS is formulated with regard to this evidence, which includes information on cost effectiveness.

The criterion traditionally used to assess cost effectiveness is the magnitude of a statistic known as the "incremental cost effectiveness ratio," defined as the difference in costs between two technologies divided by the difference in their benefits.<sup>5</sup> The lower the incremental cost effectiveness ratio, the more cost effective a technology.

We evaluated the association between source of funding and the results from economic evaluations submitted to NICE's technology appraisals programme.

### Methods

Detailed information regarding the technology appraisals programme is available elsewhere ([www.nice.org.uk](http://www.nice.org.uk)). Briefly, after receiving the final scope of the appraisal (the scope document outlines the patients, interventions, comparator technology, and outcomes of interest) from NICE, consultees have about 15 weeks to submit relevant clinical and economic evidence. During, and for a short period after, this time, the assessment group also undertakes its independent assessment of the clinical and economic evidence.

The appraisals committee meets twice to discuss the evidence. After the first meeting, the committee's preliminary recommendations are outlined in a document called the appraisal consultation document. At the second meeting, consultees' comments on this document are discussed and a final appraisal determination is produced. If consultees do not appeal against it, this document becomes the basis of NICE guidance.

Each manufacturer and the assessment group typically submit an economic evaluation so that the resulting incremental cost effectiveness ratio estimates can be compared. The institute has published its preferred rules regarding economic methods,<sup>6</sup> but they are not absolute.

When possible, we extracted estimated incremental cost effectiveness ratios for the first 62 technology appraisals, issued between March 2000 and May 2003. To do this, we needed to address several issues. Firstly, although 62 appraisal reports had been issued, some contained guidance on more than one aspect of a technology. Each of these recommendations was therefore treated as a separate observation for which cost effectiveness

## Categories of cost effectiveness

Category	Incremental cost effectiveness estimate per unit of health outcome
0	Negative incremental cost effectiveness ratios favouring use of technology or, in case of cost minimisation analysis, cheaper than alternative technology
1	From £1 to £15 000
2	From £15 001 to £30 000
3	More than £30 000
4	Negative incremental cost effectiveness ratios not favouring use of technology or, in case of cost minimisation analysis, more expensive than alternative technology

estimates could exist. Secondly, we included pairings of estimates in the final dataset only if the manufacturer and the assessment group had each submitted at least one incremental cost effectiveness ratio for a technology using the same health outcome measure—for example, a manufacturer's estimate of the incremental cost per quality adjusted life year (QALY) for drug A was matched to the incremental cost per QALY for the same drug as estimated by the assessment group. However, if the assessment group had expressed its incremental cost effectiveness ratio as an incremental cost per life year saved, this pairwise comparison was excluded from the analysis. We assumed that results from cost minimisation analyses indicated clinical equivalence, together with an increase or decrease in cost and included them in the analysis accordingly.

Over the course of an appraisal, but before the final appraisal determination, manufacturers and assessment groups may revise their original incremental cost effectiveness ratio estimates. However, we restricted our analysis to estimates contained in first submissions.

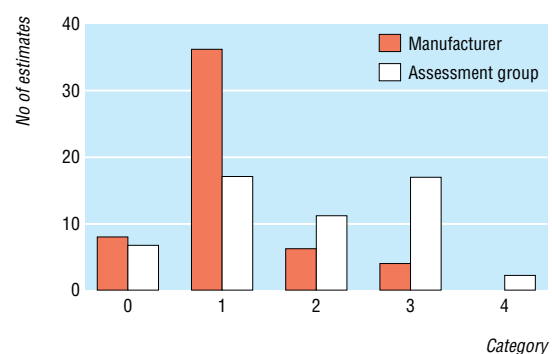
We abstracted estimates from published guidance, assessment reports, and manufacturers' submissions, and these were double checked by a second reviewer. In the small number of instances where the two reviewers' estimates did not match, we sought opinion from the relevant technology analyst at NICE.

### Statistical analysis

We used two different approaches to analysis. The first categorised incremental cost effectiveness ratio estimates according to the scale shown in the table. Although this scale does not reflect any rating system used by the institute at the time these analyses were submitted, we chose the value of £30 000 (€42 792, \$55 767) per unit of health outcome as one upper limit because of published comments on NICE's decisions.<sup>7–8</sup> By comparing categories of estimates (as opposed to continuous variables) we avoided some of the problems posed by negative estimates.<sup>9–10</sup>

Our null hypothesis was that the incremental cost effectiveness ratios produced by the assessment groups would not differ systematically from those produced by manufacturers. As the data were not continuously distributed, we performed pairwise comparison using a two tailed Wilcoxon signed rank test.

Our second analysis made only one comparison between the estimates of the assessment groups and those of the manufacturers involved in each technology appraisal, because more than one pairwise comparison of incremental cost effectiveness ratio estimates could occur (as in the first method). For example, there might have been more than one manufacturer per appraisal or separate results for different subgroups of patients. Such clustered estimates will almost certainly be correlated and, if all such estimates are used, will overestimate the accuracy of the dif-



**Fig 1** Distribution of estimates of incremental cost effectiveness ratios from assessment groups and manufacturers according to category of cost effectiveness (see table)

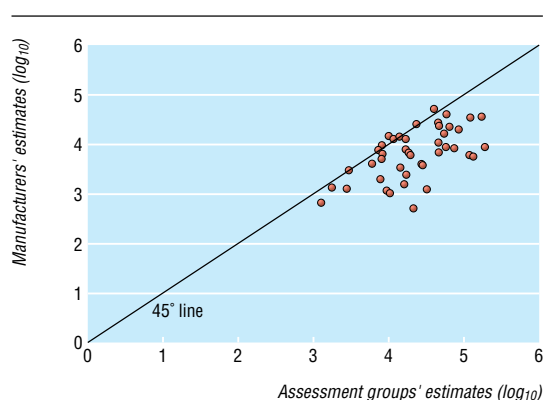
ference being investigated. Therefore, for a given technology appraisal, we compared the relevant pairwise comparisons where the manufacturer's estimate was less than the assessment group's estimate with the number of comparisons where the manufacturer's estimate was greater than the assessment group's estimate. Ties were excluded. We used the binomial distribution to test the null hypothesis that, across appraisals, the assessment group estimates would be lower than those of manufacturers as often as they would be higher. Unlike the first method, this analysis was based on point estimates rather than categories of cost effectiveness.

### Results

Of the 62 appraisals, we excluded 35 because the assessment group alone did not produce an incremental cost effectiveness ratio ( $n=20$ ), there was no manufacturer or the manufacturer did not produce an estimate ( $n=3$ ), the assessment group and the manufacturer did not produce an estimate ( $n=8$ ), or different measures of health benefit were used ( $n=4$ ). The 27 remaining appraisals contained 54 pairwise comparisons.

Over 80% of the manufacturers' incremental cost effectiveness ratios were between dominant (meaning that the appraised technology was considered to be less costly and more effective than the comparator technology) and £15 000 per unit of outcome (categories 0 and 1 combined) whereas the assessment groups' estimates were more uniformly distributed over the five categories (fig 1). Only four (7%) estimates from manufacturers were above £30 000 per unit of outcome compared with 19 (35%) estimates from the assessment groups.

There were 25 cases of negative rank (where the manufacturers' estimates were lower than those of the corresponding assessment group), 29 ties (where the estimates were in the same category), and no positive ranks ( $P < 0.01$ ; see box for example). Sensitivity analysis around the category definitions did not alter the significance of this result. Figure 2 plots the log of these pairwise comparisons; the 45° line indicates identical incremental cost effectiveness ratios. Points above this line indicate that a manufacturer's estimate was higher (less favourable) than the estimate produced by the assessment group, whereas points below this line indicate the manufacturer submitted a lower (more favourable) estimate than that of the assessment group. The analysis conducted with the second method reduced the 54 pairs of incremental cost effectiveness ratios to 27 pairs, of which two were a tie. The analysis showed that in 21 instances the manufacturers' estimates were lower than the assessment groups' estimates and in the four remaining they were higher ( $P < 0.001$ ).



**Fig 2** Logged pairwise comparison of incremental cost effectiveness ratios. Only 45/54 pairs of recommendations have been plotted because remaining nine pairs contained at least one negative estimate and could not be logged. However, in six of nine pairs estimates from the two sources were broadly similar and in three pairs estimates reported by manufacturers were much more favourable towards the technology than those reported by the assessment groups

## Discussion

We assessed whether results from economic evaluations by contracted academic centres (assessment groups) and by manufacturers of technologies were similar. Our analysis showed that the estimates of incremental cost effectiveness ratio submitted by manufacturers were on average significantly lower than those provided by the assessment groups. This finding is similar to those reported by others.<sup>2-4</sup> Unlike previous studies, however, ours is based on specific pairwise comparisons of estimates that have been put forward for the same purposes and constructed with the same terms of reference (the scope document).

Whether differences between competing cost effectiveness estimates ultimately matter in terms of decision making depends on several factors, including the level of uncertainty surrounding the incremental cost effectiveness ratios<sup>11</sup> and their absolute value. One method of assessing this would have been to divide the dataset into two, with one section including all pairings that contained at least one estimate above an implied maximum threshold of willingness to pay and another containing the remaining pairs of observations. However, the dataset was considered to be too small to do this. Nevertheless, when differences in estimated incremental cost effectiveness ratios do lead to different conclusions, the role of the appraisal committee is to judge the appropriateness of each evaluation and to determine which estimate is the most reasonable.<sup>12</sup>

NICE will soon be adopting a “reference case” approach designed to be the most appropriate for the NHS to improve consistency across all submitted economic evaluations.<sup>13</sup> It is

### Example of a pairwise comparison—appraisal No 22

#### Guidance on the use of orlistat for the treatment of obesity in adults

The manufacturer estimated the incremental cost effectiveness ratio to be about £10 500 per QALY, whereas the assessment group estimated a baseline estimate of about £46 000 per QALY. Therefore, for the first method we coded the manufacturer’s estimate as category 1 and the assessment group’s estimate as category 3, indicating a negative rank. For the second method, the single manufacturer ratio was less than that of the assessment group.

therefore feasible that differences between estimated incremental cost effectiveness ratios submitted will be smaller in the future.

We used two different analyses because, though the first method was more informative, we were concerned about the possibility of overstating the accuracy of the difference being investigated. Both approaches, however, produced similar results and led to the same conclusion.

### Limitations of study

A limitation of this study is that the evaluations submitted by a manufacturer and the corresponding assessment group are not totally independent, as the assessment group has usually had the opportunity to review the manufacturer’s evaluation before completing its own. In contrast, manufacturers do not have the opportunity to review the assessment group’s evaluation until it has been submitted to the institute. It is difficult to gauge, however, whether this is likely to lead to systematic differences in the paired ratios or influence the extent or direction of this difference.

It is also feasible that differences in estimates are not uniform across all categories of cost effectiveness. Logic would suggest that because technologies with high incremental cost effectiveness ratios are less likely to be recommended for use within the NHS, groups with vested interests in a technology might underestimate incremental cost effectiveness when the “true” estimate is high than when it is low. Insufficient data existed to test this hypothesis.

The reasons that incremental cost effectiveness ratios varied within each appraisal have not yet been investigated. For the moment, we can conclude only that estimates from manufacturers and assessment groups were significantly different from each other, not that biased results were associated with either type of organisation.

### Conclusions

In conclusion, we have shown that incremental cost effectiveness ratios submitted to NICE’s technology appraisals programme by different types of organisation were significantly different from each other. These findings undoubtedly pose questions as to the appropriate methods and processes used by the institute (and reimbursement agencies in general) when determining cost effectiveness. They also highlight the need for decision makers to have unhindered access to the methods used to produce cost effectiveness estimates so that they can be examined in detail.

We thank all the members of the appraisals team at NICE who helped collect the data, Caroline Sabin (professor of epidemiology and medical statistics, University College London) for her advice on the statistical analysis, Martin Buxton (professor of health economics, Brunel University) for his comments on earlier drafts of the manuscript, and the reviewers for their constructive comments.

**Contributors:** The study was conceived by AJF. AJF and DF presented a forerunner of this paper at an international health economics association conference. AHM and AJF were primarily responsible for further developing the study design. MG collected the data and performed the statistical analyses. AHM, AJF, and DF double checked all abstracted information. All authors contributed to the writing of the final manuscript. AHM is the guarantor.

**Funding:** None.

**Competing interests:** AHM and AJF are members of NICE’s appraisals team. DF was a member of the appraisals team when this study was undertaken.

**Ethical approval:** Not required.

- 1 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
- 2 Friedberg M, Saffran B, Stinson TJ, Nelson W, Bennett CL. Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *JAMA* 1999;15:1453-7.

### What is already known on this topic

Economic evaluations are used to produce estimates of the cost effectiveness of healthcare technologies

Profit making organisations are more likely to report favourable outcomes from clinical and economic studies than non-profit organisations

One of the key roles of NICE's appraisal committee is to judge the appropriateness of each evaluation and to determine which estimate is the most reasonable

### What this study adds

Economic evaluations submitted by manufacturers to NICE's technology appraisals programme were significantly more favourable than evaluations produced by academic research groups

The findings pose questions as to the appropriate methods and processes used by NICE (and reimbursement agencies in general) in the determination of cost effectiveness

- 3 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326: 1167-1170
- 4 Azimi NA, Welch HG. The effectiveness of cost-effectiveness analysis in containing costs. *J Gen Intern Med* 1998;10:664-9.

- 5 Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press, 1997.
- 6 NICE. Guidance for manufacturers and sponsors. [www.nice.org.uk/pdf/technicalguidanceformanufacturersandsponsors.pdf](http://www.nice.org.uk/pdf/technicalguidanceformanufacturersandsponsors.pdf) (accessed 6 May 2004).
- 7 Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 2004;13:437-52.
- 8 Raftery J. NICE: faster access to modern treatments? Analysis of guidance on health technologies. *BMJ* 2001;323:1300-3.
- 9 Briggs A. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 2000;17:479-500.
- 10 Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between strategies for gastroesophageal reflux disease. *Med Decis Making* 2002;22:290-308.
- 11 Van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E ratios alongside clinical trials. *Health Econ* 1994;3:309-19.
- 12 Garrison LP. The ISPOR good practice modelling principles—a sensible approach: be transparent, be reasonable. *Value in Health* 2003;6:6-8.
- 13 Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.

(Accepted 18 October 2004)

doi 10.1136/bmj.38285.482350.82

Health Economics Research Group, Brunel University, Uxbridge UB8 3PH  
A H Miners *visiting research fellow*

Economics Department, University of York, Heslington, York YO1 5DD  
Martina Garau *MSc student*

Health Policy Unit, Department of Public Health and Policy, London School of Health and Tropical Medicine, London WC1E 7HT  
Dogan Fidan *honorary research fellow*

Department of Community Health Sciences, St George's Hospital Medical School, London SW17 0RE  
A J Fischer *senior lecturer*

Correspondence to: A H Miners [Alec.Miners@nice.nhs.uk](mailto:Alec.Miners@nice.nhs.uk)