# Quality and economic models of cost effectiveness:

# A review of models used in key decisions by the National Institute for Clinical Excellence (NICE) in the UK

J Raftery, C Hendron, L Yao & R Taylor
Health Economics Facility, University of Birmingham.

**Health Economists Study Group meeting**
**January 2004, Paris**

**Confidential – not for quotation**

**Comments welcomed j.p.raftery@bham.ac.uk**

Address for correspondence: Professor James Raftery, Health Economics Facility, HSMC, University of Birmingham, Park House, 40 Edgbaston Park Road, Birmingham B15 2RT, United Kingdom. Email: j.p.raftery@bham.ac.uk

**Abstract**

NICE was established to provide patients, health professionals and the public with authoritative, robust, and reliable guidance on the use of health technologies within the National Health Service. In this capacity, the NICE Appraisal Committees appraise 'clinical and cost effectiveness' on the basis of stakeholder submissions and independent health technology assessments. The balance between clinical and cost effectiveness is not stated and may vary by topic. On each topic (health technology) appraised, NICE issues guidance to the NHS which is mandatory.

Both the stakeholder (mainly industry) submissions and the independent technology assessments usually include models of cost effectiveness. To meet the NICE specification of providing results in terms of incremental cost effectiveness ratios (ICERs), these models often combine the results of clinical trials (short term, surrogate outcomes) with observational data on survival and quality of life.

Standards for assessing quality in such modelling are underdeveloped. A recent systematic review of the relevant literature[1] showed broad agreement in the criteria which might help assess quality but little detail of how these might be operationalised.

Although NICE had issued guidance on 64 health technologies by August 2003, the quality of the cost effectiveness models mattered most in those where it was cited as a reason for rejecting, or restricting access to, the technology. The review reported here focuses on how modelling was used by NICE in its consideration of 18 such controversial topics, based on their high and contested ICERs.

The results indicate how criteria for assessing quality in economic models might be developed. The purpose of each model needs to be stated clearly Although a degree of transparency is ensured by having competing models available to NICE, it is not clear that non-expert NICE Appraisal Committee members, nor indeed anyone other than the experts who manipulate the models, understand the models. Sensitivity analysis is shown to be central to the process whereby NICE comes to decisions, and often requires further customised modelling. External consistency as measured by comparison between company and independent models is shown to be poor.

Training is required for the particular types of modelling required by NICE. This should arguably include examples of good practice as well as quality criteria. Making models available for further use would be helpful in such training. Since models synthesise experimental and observational data, improved methods for assessing the quality of observational data are also required.

Research priorities might include more detailed comparisons between different approaches including the usefulness to the NICE Appraisal Committee of probabilistic sensitivity analysis including Cost Effectiveness Acceptability Curves.

[1] Taylor B, Raftery, How can the quality of healthcare economic models be assessed? A review and evaluation of current guidelines. Submitted for publication.

**Introduction**

The structure of the paper is as follows:

a) methods
b) identification of a list of the most controversial topics considered by the NICE Appraisal Committee (AC) with a focus on those with a relatively high and contested ICER
c) review of decisions by NICE on each topic with particular emphasis on the role of health economic modelling
d) outline of the criteria for assessing quality in health economic modelling
e) application of these criteria to NICE's most controversial and contested topics
f) conclusions and ways forward.

Appendix 2 provides more details of the Guidance topics for those less familiar with them.

**A. Methods and Sources**

The sources of all the information on NICE's use of models, including the ICER estimates from companies and the independent academic teams are taken from published NICE reports (all of which are available on the NICE website). The key documents for each appraisal were the guidance and the TAR report.

Data were extracted from these by one author (CH) and independently checked by another (JR).

The number of guidances included was determined as those 64 published by end August 2003. However, since some appraisals dealt with more then one technology and/or patient sub-group, 103 technology/patient groups were identified.

For simplicity we have relied on the guidance number as issued by NICE for identification. These have been used as the key references and are not referenced otherwise in the text.

The definition of restricted use in describing recommendations by NICE refers to indications for treatment more restrictive than those for which the drug is licensed. In presenting and discussing incremental cost effectiveness ratios (ICERs) this paper quotes means and ranges even though NICE guidance has frequently quoted only ranges.

Details of the systematic review of quality assessment of modelling are available elsewhere.[2]

Discussion of the applicability of the quality criteria to the 18 topics is subjective, informed by the authors' experience and is offered here for discussion.

Although most of the authors have been part of the University of Birmingham team which provides independent academic input to NICE, we do not believe any conflict

[2] Taylor B, Raftery, How can the quality of healthcare economic models be assessed? A review and evaluation of current guidelines. Submitted for publication.

of interest is involved, not least as we have relied on published NICE documentation for what we say about NICE. For the record, the only appraisal of those considered in detail below was number 40, Infliximab in Crohn's Disease, in which JR was involved. The degree to which we are biased is as always for others to judge.

**B. Identification of the most controversial topics**

For each appraisal, a Technical Assessment Report (TAR) provides an independent academic review. This follows a specified protocol which includes a systematic review of the literature with an evidence hierarchy, review of sponsors' submissions and if appropriate a model of the relative cost effectiveness of the technology.

The sponsoring companies submit evidence on the clinical and cost effectiveness of their products, usually including a model on the relative cost effectiveness. The economic model has to be in the form of incremental cost per QALY, from the payor (NHS and PSS perspective) and discounted by 1.5% and 6% for benefits and costs. The modelling for the companies is sometimes done in-house, sometimes by independent agencies (which occasionally include the same teams who produce TARs).

The NICE Appraisal Committee is then faced with the task of appraising this evidence, weighing up the often contradictory evidence from the academic team and the sponsoring companies. It is at this point that the usefulness of the modelling is tested, in that the committee, assisted by the secretariat and more recently a decision support unit[3], review the modelling in detail, interrogate the academic modellers and perhaps request further work (either by the secretariat, the academics or the decision support unit).

The methodology we used for identifying the most controversial topics was based on two criteria:

- relatively high Incremental Cost Effectiveness Ratio (ICER) cited by the Appraisal Committee in making its decision,
- disagreement over the ICER between the sponsoring company(ies) and the independent technical assessment report (TAR).

The topics which were likely to be controversial were those with a relatively high incremental cost per QALY. Technologies appraised (regardless of whether recommended or not) by NICE to September 2003 were ranked by ICER and those with an ICER as accepted by NICE over £25k were reviewed in greater detail.

Operationalisation of the two criteria required tabulating the estimates of cost per QALY from all three sources (Guidance, TAR and company[4]). These are shown in Figure 1 which indicates that not all technologies appraised had cost per QALY or per LY. NICE quoted ICERs for 60 technology/patient groups out of a total of 103. The remainder either had no ICER calculated or were deemed cost saving. Many of the 60 had relatively low cost per (QA)LY. Only 22 had £/(QA)LY over £25k

[3] It was in recognition of the necessity for further modelling that NICE established the Decision Support Unit in 2003. The DSU is not discussed further here.
[4] Details of the company estimates are included in the TAR.

We chose these 22 technologies/patient groups for more detailed investigation (for details, see Table A1, Appendix 1). This list comprised the following 15 appraisals (with guidance number in brackets):

- ∉ beta interferon and glatiramer acetate in MS (32)
- ∉ infliximab in Crohn's disease (40)
- ∉ riluzole in motor neurone disease (20)
- ∉ imatinib for CML (50 and 70)
- ∉ orlistat for obesity (22)
- ∉ vinorelbine for breast cancer (54)
- ∉ zanamivir/oseltamivir/amantadine for influenza (15, 58, 67)
- ∉ somatropin for growth deficiency in adults (64)
- ∉ temozolomide for brain cancer (23)
- ∉ insulin glargine for diabetes (53)
- ∉ rheumatoid arthritis: etanercept and infliximab (36)
- ∉ paclitaxel for breast cancer (3 and 55)
- ∉ laparoscopic surgery for inguinal hernia (18)
- ∉ advanced colorectal cancer: irinotecan, oxaliplatin and raltitrexed (33)
- ∉ arrhythmias – implantable cardiac defibrillators (11)

Appendix 2 provides background details of these appraisals.

Since differences between the TAR and company estimates constituted our second criterion, four technologies/patient groups were rejected at this stage. These include laparoscopic surgery for inguinal hernia, which had a cost per QALY estimated in the guidance of £50k – this was deemed non controversial on the second criterion, that is no competing estimates, in turn due to lack of a company sponsor. Three technologies were rejected by NICE due to lack of evidence on effectiveness: temozolomide for brain cancer, raltitrexed for advanced ovarian cancer and amantadine for influenza in healthy adults. These 4 topics are removed leaving 18 disease/technologies, each of which goes forward to be reviewed in terms of the health economic modelling. These 18 are shown in Figure 2 and in Table 1.

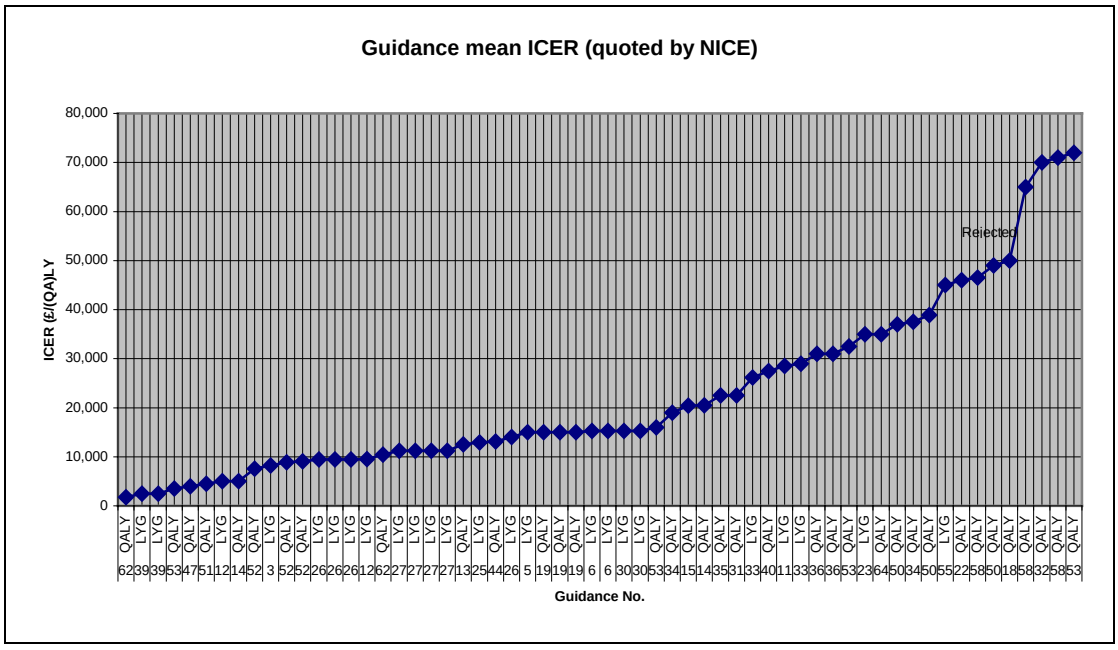


**Figure 1**
**Note – Of last 8 topics to right, starting with Number 22 all but number 55 were rejected by NICE. All others were accepted.**

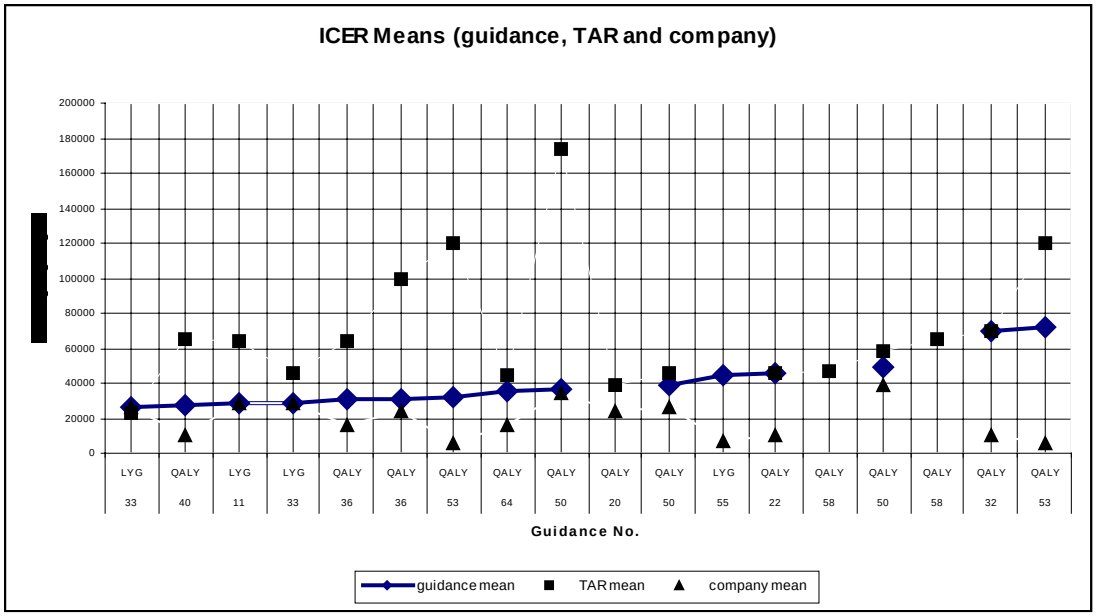


**Figure 2**
**Notes:** No company ICERs were available for Number 58 (influenza treatments: zanamivir and oseltamivir). NICE rejected the TAR estimates for these drugs as optimistic but did not quote its own estimates. No TAR ICER is available for Guidance 55, paclitaxel in ovarian cancer.

## C. Role of health economic modelling in controversial topics

This section discusses the role of health economic modelling in each of the selected controversial technologies. This is summarised in Table 2 with more details of each technology provided in Appendices 1 and 2.

### Table 1
### Summary of role of modelling in the NICE appraisal of selected controversial technologies

| | £/patient year | Key differences between company and independent models | Reaction of Appraisal Committee |
|---|---|---|---|
| **Multiple Sclerosis**<br>Beta interferon/glatiramer acetate (32) | £7-12k | Company model projected gains for lifetimes. Independent models explored shorter periods. | AC rejected company model. Against extrapolation of benefits over very long periods. |
| **Motor neurone disease**<br>Riluzole (20) | £3.7k | Independent model disputed company projections and identified a key mistake. | AC recommended drug without stating ICER. |
| **Crohn's disease (episodic)**<br>Infliximab (40) | £1.5k | Projected gains for patients' lifetimes. Independent model limited these gains to 1 year. | AC rejected company model. Inserted alternative QoL values into TAR model. |
| **Chronic Myeloid Leukaemia (CML), chronic phase**<br>Imatinib (50) | £19k-28k | A key issue had to do with the comparator: HU (low cost) or IFN alpha (high cost). TAR and company models relative to HU. | AC pro comparison as 2nd line, post IFN alpha, as per then license. ICER £37k. Recommendation widened in Guidance 70. |
| **Growth deficiency in adults**<br>Somatropin (64) | £3.3k | 2 company models and TAR model generated utility scores indirectly via QoL-AGHDA scores. | AC took view that 7 point change in QoL-AGHDA score would be necessary 'to achieve an acceptable level of cost effectiveness….in the region of £25k to 45k per QALY'. No model quoted. |
| **Obesity**<br>Orlistat (22) | £0.6k | Company models considered by NICE 'over optimistic'. No TAR model. | AC 'to attain a sufficient level of cost effectiveness in the range…£20-30k, people treated with orlistat would have to lose about 5% of body mass in each 3 months treatment'. This was consistent with national drug prescribing guidelines rather than based on modelling. |
| **Diabetes Type II**<br>Insulin glargine (53) | £0.4k | Company model (CIC) criticised as over optimistic by independent report. TAR ICER £32k Type I and £120k Type II. | Results sensitive to utility of 'fear of hypoglycaemia' and number of episodes avoided. Lower frequency of these episodes in Type II raises ICER. |
| **Rheumatoid arthritis**<br>Etanercept and<br>Infliximab (36) | £8.6k<br>£10.8k | Two company models with favourable ICERs, based on range of assumptions. Independent model had high uncertainty re key variables. | AC 'did not accept the ICERs generated by the independent model as true estimates'. It also rejected company model assumptions but incorporated its own assumptions to get the stated ICER. |
| **Influenza in healthy adults**<br>Zanamivir (15 & 58)<br>Oseltamivir (58) | £24<br>£18<br>(one-off) | Company models: cost per symptom free day. Key factors to do with a) reduced hospitalisation and death, b) numbers consulting GPs were drugs available. TAR: highly complex probabilistic model with ICERs £13k-£31k. | AC considered key assumptions in TAR model as too optimistic. Preferred own estimates. |
| **Ovarian cancer**<br>Paclitaxel (3 & 55) | £7k | Company model but no independent model. Sensitivity analyses by NICE using that model showed ICON3 results generated ICER = £45k. | First guidance pro paclitaxel in combination with platinum based therapy for patients with ovarian cancer following surgery. Revised guidance recommended it as one option, based on consideration of results of ICON3 trial. |
| **Advanced colorectal cancer** (33)<br>Irinotecan<br><br><br>Oxaliplatin | £0.4-0.8k<br><br><br>£0.5k | Independent model put ICER for 1st line at £29k/PFLY and for 2nd line, £23k (17k-28k) per LY gained. Meta analysis by company had error. | AC recommended use in second line. It considered that £/QALY would be 'significantly higher' than £29k/LY due to the quality of life of those with advanced colorectal cancer.<br>Oxaliplatin – as irinotecan except as option for 1st line in patients with metastases confined to liver which may become resectable following treatment. |
| **Arrhythmias** – implantable cardiac defibrillators (11) | £22k<br>(one off) | The company model was commercial in confidence. No independent model was produced. | The AC accepted the company model. |

The high cost (Table 2) of many of the technologies appraised is noteworthy, particularly as most (all except 11 and 58) have to be taken continuously.

Of the 18 technology/patient groups considered, NICE recommended against use in 7:
50- Imatinib in accelerated CML: ICER: £39k (£22k-£56k) and blast phase: ICER £49k (££33k-£65k)
22 – Orlistat in obesity (use only in very restricted group): ICER £46k (£19k-£55k)
58 – influenza in healthy adults: oseltamivir: ICER £47k (£18- £75k) and zanamivir: ICER £65k (£30-£100k)
32- beta interferon/glatiramer acetate in multiple sclerosis: ICER £70k (£35-£105k)
53-insulin glargine in type 2 diabetes: ICER £120k.
Of the other 11, NICE recommended use in the specified patient sub-groups.

Over time, the ICERs which NICE has accepted appear to have risen to just under £40k, for example with imatinib for chronic CML (ICER £37k, range £36-£38k) and riluzole (mean £39k, range £34-£44k). In general those technologies shown in Figure 1 with ICERs above £40k have been rejected. The one exception, recommended with an ICER of £45k, was paclitaxel in ovarian cancer. This drug had been recommended in Guidance 3 but revised in guidance 55 following the report of an ongoing trial. Insertion of the results of that trial in the company model (no independent model available) yielded an ICER of £45k, which was noted in the Guidance. Having recommended use of paclitaxel in 1999, NICE chose not to recommend against its use in 2003, but rather to recommend it as one option.

A benchmark was arguably set by NICE's verdict on beta interferon/glatiramer acetate in multiple sclerosis. These drugs (treated as one above) were rejected mainly on the evidence of high ICER (mean £70k) after repeated modelling and data collection.

Imatinib for CML has been notably difficult, not least in choice of comparator (HU or IFN alpha). In guidance 50 NICE recommended in its favour for chronic disease (ICER £37k, range £36k-£38k) but against in accelerated (ICER £39k, range £22k-£56k) and blast phase (ICER £65k, range £30k-£100k). Like paclitaxel in ovarian cancer, this is a life extending treatment[5].

Several positive recommendations were made conditional on patients' progress (obesity, growth deficiency). Conditions also applied by Government in its funding of beta interferon/glatiramer acetate for MS. The setting of conditions was based on modelling in two out of these three instances. Setting conditions is applicable to treatments which improve Quality of Life but not to mortality gains. Mortality gains (imatinib, paclitaxel) required survival analysis but were difficult to model due to short duration of trials and lack of long term data on controls.

In some instances the means by which the Appraisal Committee estimated ICERs which it quoted in support of its decision were unusual. With infliximab in Crohn's

[5] In a subsequent guidance number 70 in late 2003, NICE changed its recommendations to include imatinib as first line treatment in chronic CML and also permitted its use in both accelerated and blast phase. The rationale for this was that IFN alpha was the relevant comparator despite having an ICER relative to HU of over £1m.

disease, it used very incomplete data on Quality of Life from a relevant group of patients in an ongoing trial (placebo arm)[6] to derive an ICER of £29k compared to the TAR report later modelling of £65k.

Modelling appears to have been central to these decisions, which generally cited high ICERs as a major part of the rationale for the decisions made. NICE appears to have often relied heavily on independent models in making some decisions – examples include 32 (beta interferon/glatiramer acetate in multiple sclerosis), 40 (infliximab in Crohn's disease), 50 (imatinib in CML), 64 (somatropin for growth deficiency in adults) and 53 (insulin glargine in diabetes).

Wide discrepancies existed between the ICERs from companies (low), TARs (high) with NICE generally citing an intermediate level. The company models were often criticised for over optimism, sometimes by NICE, sometimes by TAR. Since details of company models are often Commercial in Confidence (CIC) and hence not reported in the TAR report, it is difficult to establish the sources of their low estimates.

Not all TARs submitted a model (11 - arrhythmias and ICDs, 3 & 55 - paclitaxel in ovarian cancer, 22 - orlistat in obesity did not). When they did not, this left the Appraisal Committee in a difficult position, particularly since these were controversial topics. In one of these (11, arrhythmias and ICDs) it accepted the company model in full. It is not a contractual requirement for each TAR team to provide an independent model, and it is not always apparent until late in the process whether an independent model is required.

NICE several times rejected the independent model in whole or part: 20 - riluzole in motor neurone disease; 36 - etanercept and infliximab in rheumatoid arthritis; and 58 - influenza drugs oseltamivir and zanamivir. The latter—58—is interesting in that it was the only model among the 18 topics reviewed which included cost effectiveness acceptability curves.

[6] These data have not been published nor made available to the authors despite requests.

## D. Quality criteria for modelling

The results of a systematic review of modelling and quality are shown in Table 3. The ten criteria shown were identified in 10 reviews[7]. The frequency with which each criterion was proposed is also shown in Table 3.

**Table 3**
**Criteria identified in systematic review of modelling with frequency each criterion was noted (total papers = 10)**

| Criteria | Frequency (from 10 papers) |
|---|---|
| **Transparency** | 10 |
| **External Consistency** | 9 |
| **Incorporation of Model Parameter Values** | 9 |
| **Appropriateness & Identification of Model Parameter Values** | 9 |
| **Sensitivity Analyses** | 8 |
| **Internal Consistency** | 8 |
| **Disease Aetiology & Disease Management / Appropriateness of Treatment** | 6 |
| **Time Horizon** | 5 |
| **Feasibility of Options** | 5 |
| **Perspective of model** | 5 |
| **Cycle Length** | 4 |
| **Purpose of model** | 3 |
| **Summary /10** | |

[7]The papers were as follows:
Weinstein, M.C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C., Luce, B.R. Principles of good practice for decision analytic modelling in health care evaluation: report of the ISPOR Task Force on good research practices – modelling studies. *Value in Health*. 2003, **6**(1): 9-17.
Nuijten, M.EJ.C. The selection of data sources for use in modelling studies. *Pharmacoeconomics.* 1998, **13**(3):305-16.
Buxton, M.J., Drummond, M.F., Van Hout, B.A., Prince, R.L., Sheldon, T.A., Szucs, T., Vray, M. Modelling in economic evaluation: an unavoidable fact of life. *Health Economics.* 1997 **6**:217-27.
Weinstein, M.C., Toy, E.L., Sandberg, E.A., Neumann, P.J., Evans, J.S., Kuntz, K.M., Graham, J.D., Hammitt, J.K. Modeling for health care and other policy decisions: uses, roles and validity. *Pharmacoeconomics.* 2001, **4**(5): 348-61.
Eddy, D.M. Technology assessment: the role of mathematical modelling. In, Mosteller, F. (Ed) *Assessing medical technologies.* Washington DC: National Academy Press. 1985 pp.144-53.
Halpern, M.T., Luce, B.R., Brown, R.E., et al. Health and economic outcomes modelling practices: a suggested framework. *Value in Health.* 1998 **1**:131-47.
McCabe, C. , Dixon, S. Testing the validity of cost-effectiveness models. *Pharmacoeconomics.* 2000 **17**(5): 501-13.
Soto, J. Health economic evaluations using decision analytic modelling. Principles and practices – utilisation of a checklist to their development and appraisal. *International Journal of Technology Assessment in Health Care.* 2002 **18**(1): 94-111.
Sculpher, M. Fenwick, E., Claxton, K. Assessing quality in decision analytic cost effectiveness models. A suggested framework and example of application. *Pharmacoeconomics* 2003 **17**(5):461-477.Sonneberg, F.A., Roberts, M.A., Tsevat, J. Toward a peer review process for medical decision analysis models. *Medical Care*. 1994 **32** Suppl.:JS52-64.

**E. Application of quality criteria to NICE's controversial technology appraisals**
A summary account of the application of each of the 10 criteria to ways in which the modelling of the 18 'controversial' NICE technology appraisals is provided below.

**Purpose of models**
Surprisingly, this criterion was the least often cited in the quality assurance criteria, perhaps because it was taken for granted. However, a variety of purposes can be readily listed, from 'making the best possible case' by companies, publication by academics, through to the usefulness of the modelling to the Appraisal Committee. The NICE Appraisal Committee clearly makes considerable use of economic models in making its decisions. These are models submitted by companies or developed under contract by independent academics. Relatively little use is made of published models of the technology in question. Although these are summarised in the TAR, several factors prevent them having greater impact. These include: different contexts (whether technology, comparator or country), frequent reliance on a single trial, as well as sponsorship by companies.

**Transparency**
Transparency to experts is ensured by NICE having competing models from the sponsoring companies and the academic review teams. Transparency has undoubtedly been furthered by the development of the Decision Support Unit which undertakes additional modelling for NICE (of which no details have been published). Thus those who have expertise in modelling at NICE or in the TAR groups understand the models or at least can identify aspects of particular models which are not transparent (unsupported assumptions). However it has to be noted that many of the company models are CIC, which means that only summary details can be published. More generally, it seems unlikely that any of these models (company, TAR or DSU) can be deemed transparent by experts unless they have access to working versions of the computerised models. This is the experience of the authors in dealing with company models as well as more generally.

Perhaps more importantly, it is unclear how transparent the models are to the non-experts who make up the bulk of the membership of the Appraisal Committee. In coming to decisions that the majority[8] of committee members feel comfortable with and which they believe can be defended if there is an appeal, what matters may be robustness of results from models rather than transparency. Robustness here refers to the model delivering ICERs that vary in the expected directions when parameter values are changed and which have been tested for the range of assumptions which the clinicians on the committee (rather than the generally non medical technical experts) consider reasonable. This links to sensitivity analysis below but the important point about the degree to which committee members have confidence in models is one which requires research (of the sort reported elsewhere at this conference – see Willliams, Bryan et al.)

**Sensitivity analysis**
For robust (as defined above) cost effectiveness results, much more extensive sensitivity analysis than normally practiced is required. The Appraisal Committee often requires iterative sensitivity analysis around a baseline which itself often

[8] The degree to which the Committee relies on majority verdicts with votes is unclear. Decisions are presented as the collective view of the committee.

requires considerable exploration both through modelling and expert opinion. Normal practice here refers mainly to published articles in which only minimal sensitivity analysis is the norm (not least for lack of space but also due to the lack of serious challenge, linked to lack of any use of the results in decision making). Since NICE is generally dealing with new health technologies, the uncertainty is likely to be greater than with more mature technologies. When few trials are available, estimation of cost per QALY will generally require extrapolation of benefits well beyond the time period covered by the trial (usually months rather than years). Difficulties in extrapolation apply to both intervention and comparator, but differ also depending on whether quality of life or life years are the main outcome of interest. When life years are the focus, then survival analysis is required, in turn requiring long term data, which have almost always been lacking for the relevant subgroup of patients. When quality of life is central, as in the majority of the technologies appraised to date by NICE, then data on the utilities of patients are required. Again these are seldom available since trials generally use disease specific measures which can only be mapped to utilities with heroic assumptions. Even more rare are data on the utility values of comparator patients.

**External consistency**
The one kind of external consistency considered here is that with other models.

The striking aspect of models submitted by manufacturers to NICE is their optimism, that is they consistently claim ICERs that are much more favourable to the companies product when compared to those estimates from independent groups or the Appraisal Committee. As shown in Figure 2, these differences were considerable, with company estimates all below the putative threshold of £30k and usually well below the independent estimate.

By contrast, the TAR groups generally provide ICERs well above those eventually accepted as reasonable by the Appraisal Committee (Figure 2). Why the TAR results tend to be consistently so much higher has generally to do with use of less favourable assumptions and sometimes of different methods.

**Internal Consistency**
Internal consistency has largely been concerned with detection of mistakes, which are not uncommon (See TARs 20 and 40), but detection requires very detailed attention to modelling by experts. This point is linked to transparency above. In our experience it is only by poring over the source code that proper checking for internal consistency can be assured. Simply running a model for a range of input values is an insufficient check on whether mistakes exist.

Models are often constructed to tight deadlines and are presented with minimal and non standard instructions and data input sheets. The setting of standards for the presentation and running of models could be a step forward, provided it did not impose unnecessary restrictions on the form of model or software used (both of which may vary by topic).

**Disease Aetiology & Disease Management / Appropriateness of Treatment**
This theme is arguably critical to a good model but was not explicitly addressed in any of the models considered. The necessity of understanding the aetiology of the disease becomes particularly important when projecting health gains over long periods. Without data on aetiology and disease management, it is not possible to adequately model long term health status for control groups.

Given that NICE's remit covers both guidance on technologies and guidelines for clinical services, development of comprehensive disease based models which could provide the context for more specific modelling may be worth considering. The basis of such work has been laid by Burdens of Disease (Department of Health, 1994) and currently is being extended in pilots of programme budgeting.

**Other criteria**
The issue of the feasibility of options arises mainly as regards relevant comparators. This was particularly evident with CML and imatinib (50) when the key issue had to do with whether the relevant comparator was the relatively inexpensive hydroxyurea or the relatively high cost Interferon Alpha. Multiple options are rare given that the technologies referred to NICE are generally new. Sensitivity analysis however can require that a wide range of options be available.

The perspective of the model is non-controversial due to being specified by NICE. The NICE guidelines on economic evaluation specify an NHS/Personal Social Services perspective with any other costs to be identified under equity as appropriate. No instance of non-NHS/PSS costs was noted in the 18 controversial topics.

The main issue with the appropriateness of model parameter values, and their incorporation into models, has to do with the use of systematic methods for identifying parameter values and judging their appropriateness. Few company models appear to have based their choice of parameter and values on systematic reviews. Most appear to base their models on their pivotal trials. TAR models tend by contrast to use the results of their systematic reviews for estimates of efficacy but rely on a variety of sources for other data. TAR groups often rely on the company models for data since the latter often include unpublished data.

Cycle length is generally non-controversial as long as it is appropriate. It does not appear to have been an issue in any of the topics/patient groups reviewed above.

**F. Conclusions and ways forward**
The usefulness of health economic models to NICE's decision making process has varied by technology/disease group, reflecting on one hand issues such as uncertainty and the amount of data available, and on the other, the nature of the disease/technology (mortality/morbidity). For those technology/disease topics with relatively high ICERs (poor cost effectiveness) modelling has been particularly important in identifying subgroups of patients in which particular technologies are likely to be most cost effective.

The quality criteria from the literature are of relatively little help in assessing the quality of the models submitted to the Appraisal Committee. Instead of judging

models by such criteria, NICE has ensured competition between models from companies and independent academics as a way of testing the quality of modelling.

The quality criteria do however provide useful conceptual headings around which to discuss how the quality of modelling might be improved. Of particular importance are the purpose, transparency, external and internal consistency and parameter identification. Unless models are designed to be useful in decision making, they are unlikely to be so used. Transparency of the models used by NICE is largely limited to those experts who get access to working versions. External consistency as judged by different models reaching broadly similar conclusions is poor. Company and TAR models often differ widely, sometimes for good reason (different comparators or assumptions). In coming to its decisions on controversial topics, NICE usually commissions additional modelling reflecting its informed opinion as to the most plausible and clinically relevant assumptions and scenarios. Internal consistency, it has been argued, requires experts examining models in extreme detail as only by checking the source code can errors be reliably detected and corrected.

Most of the other suggested criteria are more technical and less helpful in assessing or improving quality in general.

The differences between the company and the independent academic models are striking and could raise questions about the objectivity of both. As a result, the role of the Appraisal Committee has been closer to that of a judge in a court of law than to a scientist in the laboratory. The quality criteria for modelling apply more to science than to court. In the appraisal context, expert witnesses become important, as does sensitivity analysis linked to the assumptions that the Committee consider most reasonable.

Greater transparency may help clarify the sources of differences between different models. At present the models are transparent only to experts and company models are often commercial in confidence with the result that they cannot be fully reported by the TAR. Central to transparency is the publication of the model used by the AC in reaching a final decision. As discussed, this is often a later run on a model which is scantily reported, if at all. When such modelling has been by the academic teams, it is of course up to them to publish these models[9]. Publication of these models is important for accountability, for replication and for updating guidance when NICE reviews the topic. For this to be, possible models need to be preserved so that they can be updated. NICE's commitment to regularly reviewing its guidance provides an opportunity to build on and improve the modelling base.

[9] The publication of the model for beta interferon/glatiramer acetate in multiple sclerosis is an exception.

## Appendix 1
## Table A1

## NICE guidance topics, with cost/QALY from independent academic reviews, sponsoring company if applicable, estimate accepted by NICE (if applicable) and comments. 1999-2003 (August)

| **Disease and technology** (NICE Appraisal Number) | **TAR mean £/(QA)LY** | **Co. mean £/(QA)LY** | **NICE mean £/(QA)LY** | **Comment** |
|---|---|---|---|---|
| **Multiple Sclerosis**<br>Beta interferon/glatiramer acetate (32) | £35-104k<br>(Mean 70k) | £10k | £35-104k<br>(Mean £70k) | NICE recommends against use of these drugs. The Government then agreed to fund their use on condition patients made progress as defined by model. |
| **Motor neurone disease**<br>Riluzole (20) | £23k (£18k-29k) | £34-£43k | Not stated. | Recommended due to nature of disease and patient value of tracheotomy free time. |
| **Crohn's Disease (episodic)**<br>Infliximab (40) | £65k | £10k | £27k | AC accepted TAR framework (1 year) but used later company QoL data. |
| **Chronic Myeloid Leukaemia (Philadelphia+,CML)**<br>Imatinib (50)<br>chronic only 2nd line<br><br>accelerated<br>blast phases | <br><br>£174k( 46k-302k v hydroxurea | <br><br>£34k(£33k-35k) v hydrroxurea | <br><br>£37k(£36k-38k) post IFN alpha<br><br>£39k(22k-56k)<br>£49k(£33k-65k) | Life extending drug.<br><br><br><br>Rejected (ICER relative to IFN-alpha)<br>Rejected (ICER relative to IFN-alpha) |
| **Growth deficiency in adults**<br>Somatropin (64) | 2 reviews:<br>-1st unproven<br>-2nd £45k/QALY | £4.5k-£32k<br>£27.5k-£37.6k | £35k (25-45k) | Recommended on condition patient grows so QoL gain 7 points, |
| **Obesity**<br>Orlistat (22) | | £10.4k (£8.4k-£16k) | £20k-£30k provided 5% weight loss each 3 months | Recommended on condition patients lost 5% weight in each 3 months of treatment. AC criticised co. model |
| **Diabètes Type II**<br>Insulin glargine (53) | £120k | £4.5k-£7.0k | £32.5-£72k (re-worked by NICE) | Restricted to 'those most likely to benefit re QoL gain (fewer hypo episodes, or fewer daily injections). |
| **Rheumatoid arthritis** (36)<br>Etanercept and<br>Infliximab | <br>£64k<br>£99k | <br>£16k (14.7k-30k)<br>£24k (£17k-45k) | <br>£27k-35k | <br>AC rejected TAR model as not 'true estimates'. AC estimates new assumptions in one company model. |
| **Influenza in healthy adults**:<br>Zanamavir (15 & 58)<br>Oseltamivir (58) | ??<br>'lower than TAR' | <br>30-100k<br>18-75k | | AC considered TAR optimistic.<br>Rejected for healthy adults.<br>Rejected for healthy adults. |
| **Breast cancer**<br>Paclitaxel<br>(3)<br>(55) | <br><br>£45k or higher | <br><br>£7.1-10.8k | <br><br>£45k | Recommended as part of combination therapy post surgery or recurrent<br>Recommended in combination with platinum based compounds as first line chemotherapy, based on awaited trial (ICON3) |
| **Advanced colorectal cancer**(33)<br>Irinotecan<br><br>Oxaliplatin | <br>£23k (17k-28k)<br><br>£45k(24k-68k) | <br><br><br>£26k (£13k-39k) | <br>£29k/LY and higher £/QALY<br><br>£26k (£13k-39k | <br>Restricted to 2nd line only<br><br>Recommended as 1st line in sub-group<br>Accepted co. model |
| **Arrhythmias**<br>implantable cardiac defibrillators(11) | £64k (40-87k ) | £29k (£26k-£31k) | £29k (£26k-£31k) | AC accepted company model |
| **Topics not meeting selection criteria** | | | | |
| **Inguinal hernia**<br>laparoscopic surgery (18) | | | £55k | <br>No formal modelling |
| **Brain cancer**<br>Temozolomide (23) | <br>n.a | <br>n.a | £35k | No formal modelling – one poor trial.<br>Rejected on basis of effectiveness |
| **Advanced colorectal cancer**(33)<br>Raltitrexed | | | | <br>Rejected on effectiveness ground |
| **Influenza in healthy adults**<br>Amantadine (58) | | £13-£129k | | Rejected on basis of effectiveness |

## Appendix 2
## Summary details of selected 'controversial' NICE Technology Appraisals

### Beta interferon and glatiramer acetate in MS

Modelling was at the centre of debate and decision by the AC. Successive rounds of modelling ensured broad consensus among the academic teams. The decision not to recommend use of these drugs arguably set a benchmark: the best estimate of cost per QALY was £35k to £104k (mean £70k). This also set a further precedent in that it rejected the drugs on the basis of a range, the lower end of which had a value close to that for other technologies for which NICE had said 'yes'. In other words, the range was not used as a confidence interval.

The model developed for NICE was later used by the Government in negotiating a deal with companies.

### Riluzole in motor neurone disease

Riluzole was 'recommended for the treatment of individuals with amyotrophic lateral sclerosis (ALS) form of Motor Neurone Disease (MND)'.

Using a published company mode and 18-month trial follow-up data, the manufacturer's submission provided a base case cost per QALY of £18k to £29k for riluzole. Based on a re-analysis of this Markov model using an alternative, more conservative estimate of time-dependent probabilities, the assessment report derived a discounted cost per QALY estimate for riluzole between £34k and £43.5k (mean £39k).

The Appraisal 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 tracheotomy free survival time. With these considerations in mind, the Committee considered that the net increase in cost to the NHS of the use of riluzole in this indication was reasonable when set against the benefit, assessed as extended months of an acceptable (to patients) quality of life.'

### Infliximab in Crohn's disease

The appraisal of the use of infliximab in Crohn's disease concentrated on patients with severe active disease, with its use in patients with fistulising disease rejected. The manufacturers' estimates of the ICER for fistulising disease were between £80k and £120k. Much more attractive ICERs were estimated in a model submitted by the company which put the ICER at £6.7k for a single infusion, £10.4k for episodic treatment and £84k for maintenance treatment (defined as repeated dosing every 8 weeks in responders). The guidance noted that

> 'The manufacturer's evaluation assumed a flare rate of 10% at 2 months. Increasing the model flare rate to 20% and 50% led to an incremental cost per QALY of £20k and £55k respectively for episodic infliximab infusion' The manufacturer's evaluation also used a number of optimistic assumptions, most notably that the benefits gained due to infliximab treatment are continued over the patients' lifetime (assumed to be over 40 years).'

The economic model developed by the TAR team (University of Birmingham) evaluated the cost effectiveness of single and episodic infliximab treatment for

patients with chronic active Crohn's disease within a time frame of 1 year and using fewer health states than were used in the manufacturer's model. This evaluation estimated the incremental cost per QALY to be between £105k and £165k per QALY for single infliximab treatment and approximately £65k per QALY for episodic infliximab treatment, based on an initial treatment and three subsequent treatments.

'While accepting the general outline of the economic model developed by the University of Birmingham, the Committee sought more data to validate the accuracy of key variables such as the extent of quality of life gain, duration of response and the number of episodes of relapse in each year for which infliximab might be used. At the Committee's request, the manufacturer made available additional patient level data for up to 54 weeks in the ACCENT 1 trial, specifically from the placebo arm which comprised patients who had either received single or episodic infliximab injections. The quality of life data at week 54 were derived from only 36% of patients who were initially enrolled into the trial as placebo maintenance patients. 'After considering the estimates derived by using these data in the Birmingham model, the Committee concluded 'that on the balance of probabilities the incremental cost per QALY for episodic infliximab treatment approached £27,500 in patients with the severest form of Crohn's disease.'

**Imatinib for CML**

This has been considered twice, in Guidance 50 and 70. Only 50 is considered here. Imatinib is highly controversial because it is considered as a major breakthrough – trials were stopped at early stages due to benefits of the drug which acted on new disease pathways. Imatinib is among a few drugs (a number of cancer drugs share this characteristic) appraised by NICE which extends life.

Guidance 50: 'Imatinib is recommended as a treatment option for the management of Philadelphia chromosome positive chronic myeloid leukaemia (CML) in chronic phase in adults who are intolerant of interferon alpha (IFN-A) therapy or in whom IFN-A is deemed to have failed to control the disease.'

The manufacturer's model estimated the ICER for imatinib compared to hydroxyurea at between £33.225k and £35k for the chronic phase, between £21.k and £30.5k for the accelerated phase, and between £33.275 and £43.5k for the blast phase.

The TAR group modified the manufacturer's model and using the least favourable set of assumptions and data inputs re-calculated ICER values for imatinib compared to hydroxyurea at between £45.6k and 301.5k for the chronic phase, between £35.6k and £56k for the accelerated phase and between £52.3k and £64.75k for the blast phase.

The committee noted comparisons between the effectiveness of imatinib and other therapies. It also noted that there were problems with comparability of the patient groups in the imatinib studies versus the other studies.

**Somatropin for growth deficiency in adults**

The drug was recommended if three criteria fulfilled: that patients 'have severe GH deficiency,...have a perceived impairment of QoL,…are already receiving treatment for any other pituitary hormone deficiencies as required.' NICE considered that cost effectiveness figures would apply only to a subgroup of patients—those who had an

improvement in QoL equivalent to an absolute change in baseline QoL-AGHDA score of at least 7 points.

**Orlistat for obesity**
This was recommended only for people who have lost at least 2.5 kg by dietary control and physical activity alone in previous month and who meet body mass index criteria. Treatment was to be restricted to adults who show specified improvements, for up to three months.

'To attain a sufficient level of C-E, in the range of a £/QALY gained of between £20,000 and £30,00, people treated with orlistat have to lose about 5% of body mass for each three months that they are maintained on treatment…'. '…licensed for patients with a BMI of 30 or more, or of 28 or more in the presence of other risk factors (hypertension, Type 2 diabetes, hyperlipidaemia).' A key point in the guidance (4.2.3) is the restriction on weight loss in order for orlistat to be in the desired range of cost effectiveness. The Committee seems implicitly to accept TAR QALYs and reject the manufacturer's on the basis of over-estimation.

**Insulin glargine and diabetes**
This drug was recommended for people with Type I diabetes but restricted for those with Type II except for those who required help with insulin injections, or whose lifestyles were significantly restricted by recurrent symptomatic hypoglycaemic episodes, or who would otherwise need twice daily basal insulin injections.

**Rheumatoid arthritis: etanercept and infliximab**
Etanercept: 'is licensed for the treatment of active rheumatoid arthritis in adults when the response to DMARDs, including methotrexate (unless contraindicated), has been inadequate.' For adults, infliximab (only in combination with methotrexate) is recommended as an option by NICE for 'clinically active RA' which has not responded to at least two other anti-rheumatic drugs.

**Influenza: Zanamivir (15)**
'…licensed for individuals aged 12 and over who present with symptoms typical of influenza when influenza is circulating in the community.' Licensed dose is a 5-day course. To be started within 48 hours of onset of symptoms. It is also recommended by NICE in 2000/01 for adults at risk when consultations for influenza rise above 50/week/100 000 population. Not recommended for healthy adults.

**Zanamivir and Oseltamivir (58)**
'Zanamivir and oseltamivir are not recommended by NICE for the treatment of influenza in children or adults unless they are considered to be 'at risk'.'

**Amantadine (58)**
According to the guidance, amantadine is '…licensed for the treatment of influenza A in people for whom complications could be expected' but 'not licensed for children under 10 years of age.' 'Amantadine is not recommended by NICE for the treatment of influenza. ….on the balance of probabilities, the drugs would not be cost effective for healthy adults'.

**Paclitaxel and ovarian cancer**
Recommended by NICE as part of a combination treatment for patients with ovarian cancer after surgery, and for patients with recurrent ovarian cancer (if not previously treated with paclitaxel).

Guidance 55 changed the recommendation post ICON3: it 'recommended that paclitaxel in combination with a platinum-based compound or platinum-based therapy alone' be 'offered as alternatives for first-line chemotherapy…'.

Recommendation falls within the licensed indications. It is difficult to give LYG data—the Appraisal Committee, using survival data from ICON3 trial, came up with much higher £/LYG than the manufacturer and suggested the possibility of a still higher £/LYG. 'Simply adjusting the manufacturer's analysis to the survival difference reported by ICON3 (hazard ratio of 0.96) suggests an incremental cost per life-year gained in the region of £45,000'.

**Laparoscopic surgery for inguinal hernia**
It is recommended by NICE only for recurrent and bilateral inguinal hernia. The totally extraperitoneal procedure is preferred and there is the requirement of restriction to appropriately trained teams.

Regarding the £/QALY, the guidance notes: 'The figure is likely to be considerably reduced if' if use is restricted to 'bilateral and recurrent hernia repair'.

**Irinotecan/oxaliplatin/raltitrexed in colorectal cancer (33)**
In UK, 'irinotecan is licensed for use in chemotherapy-naïve patients with advanced colorectal cancer in combination with 5FU/FA' (1st line) and 'as a single agent for second-line chemotherapy in patients who have failed an established 5FU-based regimen.'

Irinotecan used in combination with 5FU/FA is not recommended by NICE for first line therapy. It is recommended by NICE as a monotherapy when 5FU regimen has failed.

**Oxaliplatin**
'It is licensed in the UK for the first-line treatment of adults with advanced colorectal cancer in combination with 5FU/FA. Neurotoxic side effects, which include cumulative sensory peripheral neuropathy, are dose-limiting.' NICE did not recommend it (in combination with 5FU/FA regimen) for first line therapy, except where metastases are confined solely to the liver and may become resectable after treatment.

**Raltitrexed**
'Raltitrexed is licensed in the UK for the palliative treatment of advanced colorectal cancer where 5FU/FA-based regimens are either not tolerated or inappropriate'. However it is not recommended by NICE for treatment of advanced colorectal cancer. NICE states that its use for these 'patients should be confined to appropriately designed clinical studies'. 'Neither the assessment group nor the manufacturer

provided estimates for the cost-effectiveness of raltitrexed vs. BSC or other regimens not containing 5FU/FA.'

**Arrhythmias – ICDs (11)**
Recommended by NICE for primary and secondary prevention in specific patient groups. 'The Institute believes that the economic model contained in the joint industry (Eucomed) submission, based closely on the US evidence from the AVID trial, provides a reasonable and realistic indicative estimate of the cost-effectiveness of ICDs in secondary prevention'.