

# Systematic reviews and economic evaluations conducted for the National Institute for Health and Clinical Excellence in the United Kingdom: A game of two halves?

Michael F. Drummond, Cynthia P. Iglesias

*University of York*

Nicola J. Cooper

*University of Leicester*

**Background:** Decision analytic models, as used in economic evaluations, require data on several clinical parameters. The gold standard approach is to conduct a systematic review of the relevant clinical literature, although reviews of economic evaluations indicate that this is rarely done. Technology appraisals for the National Institute for Health and Clinical Excellence (NICE), which are fully funded, represent the best case scenario for the close integration of economic evaluations and systematic reviews. The objective of this study was to assess the extent to which the systematic review of the clinical literature informs the economic evaluation in NICE technology appraisals.

**Methods:** All NICE technology assessment reports (TARs) published between January 2003 and July 2006 were considered. Data were abstracted on the TAR topics, the primary measure of clinical effectiveness, the approach to pooling in the clinical review, the measure of economic benefit and the use, or non-use, of the systematic review in the economic evaluation.

**Results:** Forty-one TARs were published in the period studied, all of which contained a systematic review. Most of the economic evaluations (85 percent) were cost-utility analyses, reflecting NICE's guidelines for economic evaluation. In seventeen cases, the clinical data were not pooled in the review, owing to heterogeneity in the clinical data or the limited number of studies. In these cases, the economists used alternative approaches for estimating the key effectiveness parameter in the model. The results of the review (when pooled) were always used when the primary clinical effectiveness measure corresponded with the measure of economic benefit (e.g., survival). However, because preference-based quality of life measures are rarely included in clinical trials, the results of the systematic review were never directly used in the cost-utility analyses. Nevertheless, the outputs of the systematic review were used when the data were useful in estimating components of the quality-adjusted life-year (QALY) (e.g., the life-years gained, or the frequencies of health states to which QALYs could be assigned). Problems occurred mainly when the clinical data were not pooled, or when the measure of clinical benefit could not be converted into health states to which QALYs could be assigned.

**Conclusions:** Economic evaluations can benefit from systematic reviews of the clinical literature. However, such reviews are not a panacea for conducting a good economic evaluation. Much of the relevant data for estimating QALYs are not contained in such reviews and the chosen method for summarizing the clinical data may inhibit the assessment of economic benefit. Problems would be reduced if those undertaking the technology assessments discussed the data requirements for the economic model at an early stage.

**Keywords:** Cost-effectiveness analysis, Meta-analysis, Health policy

Economic evaluations require clinical data for the assessment of the cost-effectiveness of healthcare treatments and programs. The gold standard approach is to conduct a systematic review of the relevant clinical literature, in particular for the estimation of relative treatment effects, a key parameter in economic decision analytic models.

A recent study of technology assessments undertaken in the United Kingdom showed that systematic reviews are often used and that this is the most prevalent approach (78 percent) for the estimation of relative treatment effect (4). However, several other approaches were also used to populate economic models. Studies of the general economic evaluation literature demonstrate an even lower use of systematic reviews. For example, Hanratty et al. (7) found that only a small proportion of published economic studies used systematic reviews that would have been available at the time the study was conducted. There is therefore a risk that the estimates of cost-effectiveness could be biased.

The objective of this research was to investigate the reasons why systematic reviews are not used, by examining situations where one might expect such reviews to be undertaken. There are several reasons why those conducting economic evaluations do not use data from systematic reviews. First, such reviews may not be available. Second, some economists may be unaware of the need to use unbiased estimates of clinical effect, although this seems unlikely. Much more likely is the possibility that the economist does not have the time or resources to conduct a systematic review in situations where no reviews exist. Finally, there is the possibility that, although systematic reviews exist, they are not able to be used in the economic evaluation.

The technology assessment reports (TARs) undertaken for NICE are an excellent vehicle for studying the link between systematic reviews and economic evaluations for two main reasons. First, the contract for TARs requires that the evaluation team conducts a systematic review to provide parameter estimates for an economic evaluation (usually based on an economic model). Second, there is adequate time and funding to undertake the systematic review. Therefore, if the analysts undertaking the economic evaluation component of the technology appraisal choose not to use the systematic review, it can only be because they perceive it to be unhelpful or irrelevant.

## METHODS

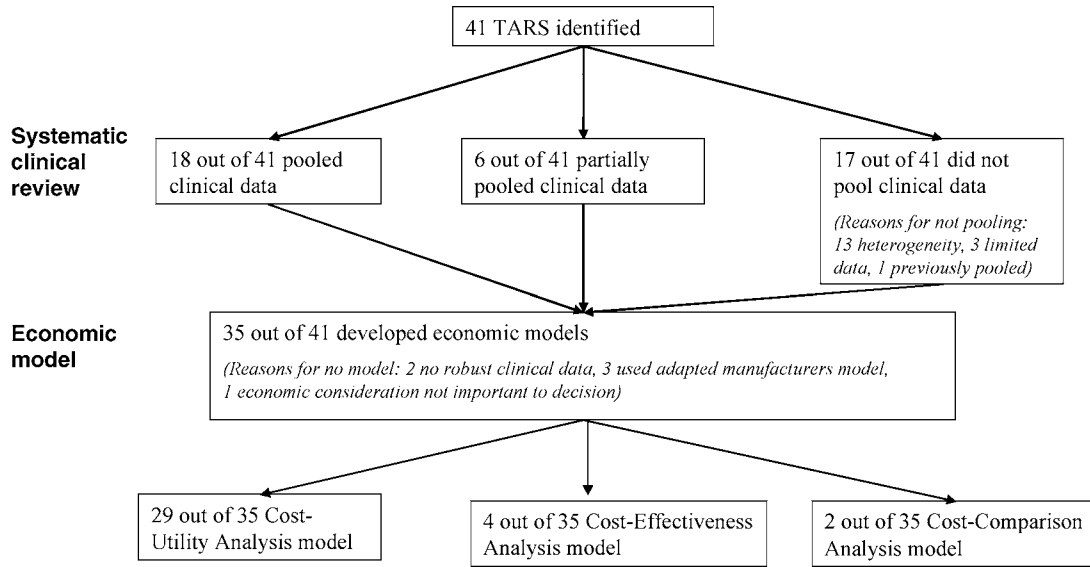
All NICE technology assessment reports (TARs) published between January 2003 and July 2006 were considered. The reports were obtained and data extracted on the following items: (i) the TAR topic and date; (ii) the identity of the evaluation team; (iii) the primary measure of clinical effectiveness and other clinical data; (iv) the approach to pooling in the (clinical) systematic review; (v) the reason given if the data were not pooled; (vi) the measure of benefit in the economic evaluation; (vii) the use, or non-use, of the evidence from the systematic review (in the economic evaluation); and (viii) the reasons given for not using the systematic review and the justification for the alternative approach to populating the economic model. These data were tabulated and analyzed in detail. The analysis concentrated on the use, or non-use, of the main measure of relative treatment effect from the systematic review, because this is the most important clinical estimate in economic evaluations.

## RESULTS

Table 1 shows the topics covered in the forty-one TARs sampled. It can be seen that a wide range of health technologies was considered. The majority were pharmaceuticals (69 percent), which reflects NICE's overall technology appraisal program.

**Table 1.** Overall Summary of TARs Considered

|   |           |
|---|-----------|
| Pharmaceuticals   | 28        |
| Drugs for ovarian cancer, diabetes, influenza, colorectal cancer, breast cancer, growth hormone deficiency, non-Hodgkin's lymphoma, bipolar disorder, rheumatoid arthritis, angina, attention deficit hyperactivity disorder, epilepsy, and so on |           |
| Devices   | 4         |
| Dual-chamber pacemakers, vaginal tape, insulin pumps, coronary stents   |           |
| Procedures  | 9         |
| Electroconvulsive therapy, patient education, cervical cancer screening, perfusion scintigraphy, cognitive behavioral therapy, fluid replacement therapy, endometrial ablation, laparoscopic surgery, computerized cognitive behavioral therapy   |           |
| <b>TOTAL</b>  | <b>41</b> |



**Figure 1.** Link between the systematic review and economic model. TARs, technology assessment reports.

In all forty-one TARs, a systematic review of the clinical literature was conducted. However, it can be seen (Figure 1) that pooling (i.e., to produce an overall estimate of clinical effect) was not possible in 17 cases, the main reason being the heterogeneity in the clinical studies. This points to the first problem encountered by the analysts producing the economic model, because the model requires estimates of the key parameters, in particular clinical effectiveness. Another issue facing the economic modeler is that the pooled statistic (e.g., odds ratio) may need some manipulation before its use in the economic model, although this should not be a major obstacle. The most useful summary measure for the modeler is the relative treatment effect, because this is usually considered to be generalizable across studies and can be applied to different baseline risks to estimate absolute effects for different patient populations.

However, despite the lack of pooling, the analysis of the TARs showed that a model was developed in thirty-five cases, suggesting that on eleven occasions, this was done in the absence of a pooled estimate of clinical effectiveness from the systematic review. In these cases, the strategies used by the economic analysts were (i) to produce a pooled estimate of clinical effectiveness from a subset of the clinical studies (i.e., the major trials), (ii) to use expert opinion, or (iii) to use pooled estimates contained in the submission of data from the manufacturer. Each of these strategies, although understandable, has the potential for introducing bias.

On six occasions, the economist did not develop a model. In two of these cases, the economic analyst concluded that there were not robust clinical data. Other reasons for not developing a model were that the manufacturer’s model could be used or adapted, or that the economic considerations were

not an important factor in choosing between the technologies under consideration.

A total of forty-two economic analyses were conducted in the TARs. (One TAR contained two analyses.) The vast majority (85 percent) were cost-utility analyses, with the measure of economic benefit being expressed in quality-adjusted life-years (QALYs). This reflects the preferences expressed in NICE’s methodological guidance to those undertaking technology appraisals (3:9). In the other cases, two TARs expressed the economic benefits in terms of life-years gained, five used a disease-specific measure (e.g., percentage of therapeutic response, invasive cancers avoided or caries cured), and three undertook cost comparisons only (on the grounds that the benefits of the technologies being concerned were broadly equivalent). On the occasions where QALYs were not estimated, the economic analysts argued that good quality of life data were lacking, that the direct clinical measures from the trials were more reliable, or that benefit measurement was not necessary.

However, thought needs to be given regarding the most appropriate statistic for use in the decision model. For example, if the clinical outcome is survival, the clinical trials, and thus the systematic review, normally report the median (the most appropriate summary statistic for clinical effectiveness), but the mean survival is the more appropriate statistic for the decision model (i.e., in economic evaluations mean cost is normally calculated; therefore, this needs to be compared with mean survival). This was an issue in the NICE appraisal on neuraminidase inhibitors (10). The clinical outcome was median time to alleviation of all symptoms, but for the model, the meta-analysis was redone using the mean time. (In cases where the estimate could not be supplied by the manufacturer, some distributional assumptions were

made). Therefore, even if a systematic review has been carried out on clinical effectiveness, it may be the wrong statistic for the economic model.

However, in this sample of reviews, NICE's preference for economic benefits to be estimated in QALYs gained was a major factor limiting the direct use of the systematic review estimates in the economic model. Health utility measures are not often included in clinical studies, so it is highly unlikely that the estimate of relative effectiveness from the review would be expressed in QALYs. However, where available, pooled estimates of survival differences from the systematic review were used by the economic analyst in calculating the "life-years gained" component of the QALY. Also, in a few cases, estimates from the systematic review were used in estimating the quality of life (or utility) component of the QALY. An example of this was in TAR No. 64 on human growth hormone in growth hormone-deficiency adults. Here, the review of the clinical evidence was focused on non-preference-based health-related quality of life instruments (Nottingham Health Profile and the Quality of Life Adult Growth Hormone Deficiency Assessment). Consequently, a regression model, developed as part of the industry submission, was the only vehicle available to map Quality of Life Adult Growth Hormone Deficiency Assessment onto a preference-based index for use in the economic analysis (3).

On the other hand, in some situations, the pooled estimate of clinical effectiveness from the systematic review was not very helpful to the economic analyst seeking to estimate QALYs gained. An example of this arose in the TAR concerning drugs for the treatment of attention deficit hyperactivity disorder (ADHD) (8). Those undertaking the systematic review decided to produce a pooled estimate of "points improvement on the Connors Hyperactivity Scale," on the grounds that this outcome was reported in the vast majority of studies and that it was a clearly defined measure. The main alternative outcome, a clinical assessment of "response," was considered to be inconsistently defined across studies. However, from the economist's perspective, a change in Connors Points Score is not easily converted into a QALY. Therefore, in constructing the economic model, an estimate of "full or partial clinical response" was obtained from a subset of the clinical studies.

## DISCUSSION

This study shows that, where adequate resources are made available for the technology assessment, economists can use the outputs of systematic reviews to provide parameter estimates for their economic models. However, three main problems remain, which can cause technology assessments to be a "game of two halves." First, in situations like that prevailing in the United Kingdom, where QALYs are preferred by decision makers, it is unlikely that these will be a direct output of the review. However, the outputs of the review are often used *indirectly* in the estimation of QALYs gained.

Second, those undertaking the systematic review may believe that it is unwise to produce a pooled estimate of clinical effect, usually because of heterogeneity in the clinical studies. This presents the economists with a dilemma, because pooled estimates are required to parameterize the model. However, even if the economists decides to produce parameter estimates by another route, they may still benefit from the thorough literature search that is a component of the systematic review, or from the narrative summary that is produced.

Third, the outcomes chosen for pooling in the systematic review may not easily lend themselves to the estimation of QALYs.

Several steps could be taken to resolve these problems. First, data instruments [such as the EQ-5D (5) or Health Utilities Index (6)] could be used in clinical trials to provide direct estimates of health utility gains. Although this would add to the cost of clinical trials, there are signs that inclusion of such instruments is becoming more common (1). Also, the SF-36 (a generic profile measure of health-related quality of life) is more regularly included in clinical trials. An algorithm then can be used to convert SF-36 data into a health index, the SF-6D (2).

Second, there should be more debate about the pros and cons of pooling the data on clinical effectiveness in different situations. Whereas pooled estimates may be problematic where there is heterogeneity in the clinical studies, the alternative approaches (e.g., using a subset of studies, or using expert opinion) may be even more unsatisfactory. In situations where a technology assessment has been commissioned, a decision usually results (even if the decision is to do nothing). Therefore, economic analysts typically believe that it is their duty to produce the best possible estimates of clinical and cost-effectiveness, even if these are subject to considerable uncertainty. This approach may be at odds with that usually followed by those undertaking systematic reviews, for whom homogeneity between studies is the main requirement to enable a quantitative synthesis of the evidence.

Third, there should be more discussion, before the systematic review, on the data to be extracted and the ways in which they will be summarized. This will maximize the chances of the estimates produced being useful for the economic model, although it has to be recognized that those undertaking the systematic reviews are ultimately restricted by the data collected in the clinical studies themselves.

## CONCLUSIONS

Economic evaluations can benefit from systematic reviews of the clinical literature. However, such reviews are not a panacea for conducting a good economic evaluation. Much of the relevant data for estimating QALYs are not contained in such reviews, and the chosen method for summarizing the clinical data may inhibit the assessment of economic benefit. Problems would be reduced if those undertaking

technology assessments discussed the data requirements for the economic model at an early stage.

### CONTACT INFORMATION

**Michael F. Drummond**, MPhil (md18@york.ac.uk), Professor, Centre for Health Economics, **Cynthia P. Iglesias**, PhD (cpiu@york.ac.uk), Senior Research Fellow, Centre for Health Economics and Department of Health Sciences, University of York, Heslington, York, YO10 5DD, UK

**Nicola J. Cooper**, PhD (njc21@le.ac.uk), Senior Research Fellow / MRC Fellow, Department of Health Sciences, University of Leicester, 2nd Floor (Room 208), Adrian Building, University Road, Leicester LE1 7RH, UK

### REFERENCES

1. Barbieri M, Drummond MF, Puig Junoy J, Casado Gomez MA, Blasco Segura PB, Poveda Andres JL. A critical appraisal of pharmacoeconomic studies comparing TNF alpha antagonists for the rheumatoid arthritis treatment. *Expert Rev Pharmacoecon Outcomes Res.* 2007;7:613–626.
2. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ.* 2002;21:271–292.
3. Bryant J, Loveman E, Chase D, et al. Clinical effectiveness and cost-effectiveness of growth hormone in adults in relation to

- impact on quality of life: A systematic review and economic evaluation. *Health Technol Assess.* 2002;6:1–106.
4. Cooper N, Coyle D, Abrams K, Mugford M, Sutton A. Use of evidence in decision models: An appraisal of health technology assessments in the UK since 1997. *J Health Serv Res Policy.* 2005;10:245–250.
5. EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy.* 1990;16:199–208.
6. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single attribute functions for the Health Utilities Index Mark 3 System. *Med Care.* 2002;40:113–128.
7. Hanratty B, Craig D, Nixon J, Rice S, Christie J, Drummond MF. Are the best available clinical effectiveness data used in economic evaluations of drug therapies? *J Health Serv Res Policy.* 2007;12:138–141.
8. King S, Griffin S, Hodges Z, et al. A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, clexamfetamine and amoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents. *Health Technol Assess.* 2006;10:1–62.
9. National Institute for Clinical Excellence. *Guide to the methods of technology appraisal.* London: NICE; 2004.
10. Turner D, Wailoo A, Nicholson K, Cooper N, Sutton A, Abrams K. Systematic review and economic decision modelling for the prevention and treatment of influenza A and B. *Health Technol Assess.* 2003;7:1–182.