

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/23733986>

# Modelling in Health Economic Evaluation: What is its Place? What is its Value?

Article in *Pharmacoeconomics* · June 2000

DOI: 10.2165/00019053-200017050-00004 · Source: RePEc

CITATIONS

177

READS

1,741

2 authors:



**Alan Brennan**

The University of Sheffield

351 PUBLICATIONS 8,723 CITATIONS

[SEE PROFILE](#)



**Ron Akehurst**

The University of Sheffield

143 PUBLICATIONS 3,983 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



The Relative Effectiveness of Pumps Over MDI and Structured Education (REPOSE) [View project](#)



EVPI/EVSI [View project](#)

# Modelling in Health Economic Evaluation

## What is its Place? What is its Value?

*Alan Brennan and Ron Akehurst*

School of Health and Related Research (SchARR), University of Sheffield, Sheffield, England

### Contents

Abstract	445
1. Health Technology Assessment and Reimbursement	446
2. Modelling	446
3. Modelling or Trials: A Redundant Debate	447
4. Reported Prevalence of Modelling	448
5. The Roles of Modelling in the Economic Evaluation of Health Technologies	448
5.1 Extending Results from a Single Trial	449
5.2 Combining Sources of Evidence	449
5.2.1 Intermediate Clinical End-Points and Final Outcomes	449
5.2.2 Extending to Relevant Comparators	450
5.2.3 Modelling Broader Systems	451
5.3 Generalising Results from One Specific Context to Others	451
5.3.1 Trials Into Practice	451
5.3.2 Generalising Between Locations	451
5.4 Modelling Prior to Trials and Studies	452
5.5 Sensitivity Analysis and Modelling Uncertainty	453
5.5.1 Type 1: Simple Sensitivity Analysis	454
5.5.2 Type 2: Threshold Analysis	454
5.5.3 Type 3: Extremes Analysis	454
5.5.4 Type 4: Probabilistic Monte Carlo Sensitivity Analysis	454
5.5.5 Type 5: Structural Sensitivity Analysis	454
5.6 Further Aspects of the Value of Modelling	455
5.6.1 Value as Communication Tools	455
5.6.2 Value in Problem Structuring – Conceptual Modelling and Subjective Information	455
5.6.3 Informing the Problem Situation when Hard Data are Impossible to Obtain	455
6. Discussion	456
6.1 The Need for Quality Assurance and Validation	456
6.2 Validation	456
7. Conclusions	457

### Abstract

This paper itemises the current and developing roles of modelling in health economic evaluation and discusses its value in each role.

We begin by noting the emptiness of the dichotomy that some commentators have sought to create between modelling and controlled trials as mechanisms for informing decisions. Both are necessary parts of the armoury. Recent literature discussions are examined and the accelerating prevalence of modelling is reported.

The identified roles include: extrapolating outcomes to the longer term; adjusting for prognostic factors in trials; translating from intermediate to final outcomes; extending analysis to the relevant comparators; generalising from specific trial populations to the full target group for an intervention and to other settings and countries; systematic sensitivity analyses; and the use of modelling for the design and prioritisation of future trials.

Roles are illustrated with 20 recent examples, mostly from within our own work analysing new or contentious interventions for the Trent Development and Evaluation Committee, which is planned to be incorporated into the UK National Institute for Clinical Excellence (NICE). Each role discussed has been essential at some point in this policy-making forum.

Finally, the importance of quality assurance, critical review and validity testing is reiterated and there are some observations on processes to ensure probity and quality.

---

This paper itemises the current and developing roles of modelling in health economic evaluation and discusses its value in each role.

## 1. Health Technology Assessment and Reimbursement

The world of health technology assessment is developing at speed. Governmental desire to control burgeoning health budgets, together with developments in health services research methodology and practice, have combined to produce much more questioning of the cost effectiveness of new and existing interventions. The focus has been particularly sharp in the case of pharmaceuticals. The Netherlands, Norway, Portugal, Spain, the UK and the USA all have guidelines for evaluation of pharmaceuticals in various stages of development, and statutory processes are already in place in Canada and Australia.<sup>[1,2]</sup>

In the UK, the recently created National Institute for Clinical Excellence (NICE) is required to appraise new technologies, particularly pharmaceuticals. NICE will build upon the work of the existing regional Development Evaluation Committees (DECs) in providing advice to Health Authorities. The Trent DEC has been serviced by the School of Health and Related Research (SchHARR) at the University of Sheffield. The authors of this paper have been heavily involved in the process of

evidence review, synthesis of data and modelling necessary to provide policy advice.

## 2. Modelling

Modelling represents the real world with a series of numbers and mathematical and statistical relationships. In the twentieth century its use has exploded, with almost universal application, from atomic physics and weather forecasting to military strategy and international business. In the health field, modelling is often used for planning budgets, workforce and location of facilities. Certain clinical fields are fundamentally based on mathematical modelling approaches, for example pharmacokinetics and epidemiology.

The literature gives some excellent introductions to modelling in broader health technology assessment.<sup>[3-6]</sup> The progression of the disease or the 'patient pathways' must be examined, how these would change given the intervention, and the subsequent costs and outcomes. Such assessments apply not only to treatment technologies but also to broader systems with more 'knock-on' consequences – health promotion, preventive and vaccination programmes, screening policies and diagnostic services. Technically, modelling frameworks include formalised approaches such as decision trees, Markov analysis, discrete event simulation and systems dynamics. In practice, models vary in complexity and investment required. The choice of

approach depends on the structure of the disease, the impact of the technology and the availability of data for its assessment. No single framework is always applicable and an increasing number of models are simply extended spreadsheet calculations.

### 3. Modelling or Trials: A Redundant Debate

Health technology assessment (HTA) has 2 distinct phases:

- gathering evidence – from randomised controlled trials (RCTs), observational studies, case control studies etc.
- processing evidence – to estimate the performance of the technology in the circumstances of interest, often circumstances that either have not or cannot be observed in a trial situation. It is in this second phase of HTA that most applications of modelling occur.

The literature discussions on the roles of modelling often take an adversarial, trials versus modelling, perspective. Luce<sup>[6]</sup> reviewed the past 30 years of cost-effectiveness analysis with particular reference to modelling. He examined the genesis of the recent policy of the *New England Journal of Medicine*<sup>[7]</sup> and the Task Force on Principles For Economic Analysis of Healthcare Technology<sup>[8]</sup> on industry funded modelling. These developments were prompted by concerns about bias and validity, which are expressed globally but most forcibly in the US and by the Food and Drug Administration (FDA).

Luce analysed this debate as a clash between 2 cultures. Biomedical researchers and the FDA have a paradigm of RCTs, experimental data and hypothesis testing. In contrast, the health economics and health technology assessment communities have a different paradigm of cost effectiveness and the need to support policy decisions. The latter recognise the ‘necessity of various types of analytical models to enrich and broaden results from experimental research when it is available and to find substitutes for experimental data when it is not available’.<sup>[6]</sup>

The advantages of RCTs are well understood. In particular, the methodology ensures that the effect

is attributable to the intervention alone through the exclusion of potential biasing factors, such as patient selection, physician suggestion and the placebo effect. RCTs meet the criteria for the best scientific evidence: replicability, verification and falsification. The statistical apparatus also enables assessment of the uncertainty in effects. However, there are problems with the direct use of RCT evidence for policy-making, and these are well reviewed by Rittenhouse:<sup>[5]</sup>

- choice of comparison therapy
- protocol-driven costs and outcomes
- artificial environment
- intermediate versus final outcomes
- inadequate patient follow-up
- selected patient and provider populations.

Many trials use placebo comparison for registration purposes and it is often only by modelling that the most relevant comparator or current mix of care can be assessed. Protocol-driven care within a trial can cause significantly higher levels of compliance, monitoring of safety and general care than occurs in practice.

Indeed, the Canadian guidelines suggest that ‘protocol-driven costs should be excluded if they would not occur as part of the intervention on a regular basis’.<sup>[1]</sup> RCTs often use intermediate rather than final outcomes and can have inadequate follow-up of dropouts or treatment failures. They can be biased because of the population selected (often healthier and more compliant than the real population) or because of self-selected providers of care (e.g. specialist clinicians who may be better at diagnosing produce more true positives using a diagnostic technology).

The issue of effectiveness versus efficacy studies continues to be fundamentally important in pharmacoeconomics.<sup>[9]</sup> A recent review of methodological problems confirms that these issues apply not only to the UK National Health Service (NHS) but also globally to pharmaceutical manufacturers, insurers, providers and practice guideline committees as well as policy-makers and consumers.<sup>[10]</sup> These problems are commonplace and well established and, as we shall see, they are also

reflected in our experience of supporting the Trent DEC.

On the other hand, the literature also provides some important cautions against inappropriate use of modelling.<sup>[11-13]</sup> Worries about combining evidence from incompatible studies, extrapolating to longer term outcomes and partial or misleading sensitivity analyses are to the fore. There is also a debate about whether and when large pragmatic trials should be commissioned rather than relying on modelling for policy decisions. Deciding when and where to invest in such studies requires an attempt to review and synthesise the available evidence and understand the remaining uncertainty in our knowledge. This assessment of the scale of uncertainty and its effect is necessary to justify the priority and inform the design of the study. In other words we need a model.

For NICE and other such bodies internationally, the reality is that modelling has a key role. Modelling and the conduct of trials are not alternatives if the intention is to give policy advice. Both form part of the necessary armoury. Certainly, the blanket rejection of the use of modelling in economic evaluation in favour of trials, as some authors have seemed to recommend, is foolish and misguided. The use of sound modelling methodology, so that 'good' models can be developed and used and 'bad' models can be told apart, is of tremendous importance as modelling plays a larger part in evaluation. The roles and examples outlined in the following discussion serve to underline the mutually important interactions between collecting basic evidence and the synthesis of data to inform policy.

#### 4. Reported Prevalence of Modelling

The Office of Health Economics database<sup>[14]</sup> summarises facets of around 3000 health economic evaluation studies between 1991 and 1997. They cover a diversity of interventions and, importantly, the vast majority involve the synthesis of evidence from non-RCT sources. There has been an enormous growth in the number of evaluations: 109 in 1991 rising to 2471 in 1997. The interventions include pharmaceutical (33%), surgical (13%), diag-

nostic (14%), screening (6%), prevention (7%), devices (6%), procedures (7%) and general care (14%).

Just 19% of the studies used RCTs to give the probability of the main clinical events, with 59% using observational studies, 3% systematic review or meta-analysis and 16% other literature review. Yet more studies used data from outside trials for the quantity of resources used – 21% used other literature review and 5% used expert judgements. Similarly, the costs of these resources, while often obtained using local costs (48%), were sometimes calculated using national publications (8%), judgement (3%) or *ad hoc* estimation from other literature (36%).

The use of formal decision analytic modelling was very limited – 2% of studies used it in the estimation of main clinical events and the resources used. Our own recent review of keywords in the NHS Economic Evaluations database also showed low levels of use of the more formal mathematical approaches – Markov (3%), simulation (3%) and decision tree (2%).

The conclusion is that the synthesis of evidence in published health economic evaluations is the norm, while fully integrated RCT-based economic evaluations are more rare. The more formal, sophisticated modelling approaches are also rarely used.

#### 5. The Roles of Modelling in the Economic Evaluation of Health Technologies

The roles identified are illustrated using examples mostly from our own very recent work for the Trent DEC, which advises Health Authorities on whether, and in what circumstances, they should fund a particular technology.

In the course of the production of 24 reports, involving perhaps 8 analyst-years of effort, it has become clear that almost every policy decision has required the combination of evidence from a variety of sources. This combination has usually required some formal modelling. In no case has the result from a clinical trial or a set of trials alone been sufficient to answer the policy question. In many cases it is impossible to imagine individual trials

that could have provided all the answers. Equally there have been examples where the absence of adequate and relevant trial information has made guidance difficult to produce and subject to wide margins of uncertainty.

The different roles and applications of modelling are identified through 5 perspectives:

- extending results from a single trial
- combining multiple sources of evidence to answer policy questions
- generalising results from one specific context to others
- modelling to inform research strategy and design
- modelling uncertainties in the knowledge base.

### 5.1 Extending Results from a Single Trial

When preparing DEC reports it is common to find clinical trials that are potentially relevant for policy, but which fall short in some key manner. Many clinical trials have end-points that may be too early, even from a clinical perspective, and it is very common for a trial to stop just when it becomes interesting to an economist pursuing data on use of resources by different groups. Modelling has value in allowing the extrapolation of shorter term data into the future by using explicit assumptions about underlying disease progression or similar outcomes.

Extending results from a single trial includes:

- extrapolating reported health outcomes to the longer term
- extrapolating the cost analysis to the longer term and discounting
- translating the clinical efficacy outcomes measured in the trial (e.g. a disease-specific scale) to long term health economic outcomes (e.g. quality-adjusted life-years)
- improving/adjusting the analysis of trial results (e.g. subgroup analysis to match the general population of interest).

Example 1: Longer Term Outcomes – Paclitaxel for Ovarian Cancer

The key trial data reported survival benefits at 4 years.<sup>[15]</sup> Statistical forecasting and scenario analysis were used to extrapolate to 7 years as

part of a sensitivity analysis. The analysis confirmed that paclitaxel was cost effective within the typical threshold for costs per life year gained.<sup>[16]</sup>

Example 2: Adjusting for Prognostic Factors – Riluzole

An analysis of prognostic factors formed an important, and contentious, element in the Trent DEC report on riluzole.<sup>[17]</sup> This was an attempt to adjust for possible bias in the original trial. The explicit nature of the modelling quantified the effect of adjustments to allow for differences in patient characteristics in the control and treatment groups, which were known independently to be important. These adjustments were crucial to the estimates of efficacy accepted by the drug licensing authorities. The analytical team was willing to accept the adjustments as valid but the DEC itself was not and recommended that the intervention should not be funded until better evidence emerges.

### 5.2 Combining Sources of Evidence

Very often other studies or data contain important evidence that can and should be utilised to inform policy. It is in this combination of data that the greatest strength of modelling, and the greatest opportunity for error, probably lies. Some forms of combining data sources are already well accepted, for example meta-analysis, itself a form of modelling. Meta-analysis carries problems common for many modelling exercises, including choice of trials to include, assessing the validity of pooled information, analysing uncertainty and balancing extending the sample size against reductions in compatibility of information.

#### **5.2.1 Intermediate Clinical End-Points and Final Outcomes**

A common reason for modelling is the desire to extend intermediate end-points to final outcomes. Buxton et al.<sup>[12]</sup> recommend that a modelled relationship between intermediate end-points and final outcomes should only be used where that relationship has been proven to exist. We wholeheartedly agree and would go further, to say that when such relationships are known, modelling them in con-

junction with trial results is highly desirable for informing policy.

For example, a proven relationship between chemical markers at 6 months after treatment and survival and quality of life at 5 years is immensely helpful both for policy decisions and for the future efficiency of clinical trials. Furthermore, the modelling can be used to calculate expected survival outcomes year by year. This is very important for NICE because it ties together appraisal of technologies, subsequent study design and audit processes.

#### Example 3: Olanzapine for Schizophrenia

The cost effectiveness of olanzapine was examined by 2 separate sets of analysts.<sup>[18]</sup> Both extended outcomes evidence based on 3-month trials to 1 or 2 years using a Markov model with transition probabilities between different disease/symptom states. The analysis required translation of intermediate outcomes to final outcomes by clinical experts. Clinical symptom scales (the British Psychological Ratings Scale and the Positive and Negative Symptom Scale) were translated into implications for patient management and resource consequences. Both analyses showed that olanzapine would be cost effective under these assumptions.<sup>[19]</sup>

#### Examples 4 and 5: Extending Intermediate Outcomes using Secondary Databases

In the treatment of patients with rheumatoid arthritis, cocktails of drugs including methotrexate and cyclosporin are currently being trialed for their 1- and 2-year impact on intermediate indicators (chemical and quality of life). As yet unpublished work is extending these short term results using a long term database on patients with rheumatoid arthritis held elsewhere.

In solid organ transplantation, a variety of new drugs have impact on acute rejection episodes in the first 6 months after transplantation. Again, as yet unpublished work is examining extension of short term acute events to subsequent long term survival using a similar long term database of patients' experience.

The development of the concepts and techniques of cross-design synthesis<sup>[20]</sup> provides another step forward in the systematic and valid combination of evidence sources. Cross-design synthesis

firstly assesses the overall quality of all kinds of studies identified. Secondly, it chooses studies based on the possibility of the elimination of individual study bias through use of complementary design. Rittenhouse<sup>[5]</sup> discusses an example of the approach: to pronounce on the generalisability of the results from an RCT, cross-design synthesis would examine the way in which patient recruitment was accomplished and compare the sample with the population of interest. It would examine the inclusion/exclusion criteria, the investigator's choices of eligible patients, the patients' willingness to participate once selected, etc. If the sample differed greatly from the target population, age or gender, linked results could be examined and a trial's results re-weighted according to the target population.

#### 5.2.2 Extending to Relevant Comparators

Estimated cost effectiveness depends on the comparator chosen as well as the studied health technology. The question of comparator is a vexed one, with debate around whether 'normal care' or 'best alternative care' should be chosen.<sup>[21]</sup> For licensing reasons, most pharmaceutical trials are done against placebo. For other technologies, the most complete studies often contain no control data at all. Models allow comparison against more than one alternative to inform policy.

#### Example 6: Comparators in *Helicobacter pylori* Eradication

A model developed by SchARR<sup>[22]</sup> studied the effects of *H. pylori* eradication on peptic ulcer prevalence and symptom-free days. Data on all existing options are incorporated for comparison with a range of conventional acid suppressant therapies. Local population, prevalence and cost data can be used. Regimens for comparison include those recommended by recent European Guidelines<sup>[23]</sup> as well as user customised options.

#### Example 1 Revisited: Paclitaxel

The paclitaxel study<sup>[16]</sup> also demanded extension to relevant comparators. Carboplatin is used in the UK, but the trial was against cisplatin and cyclophosphamide. The analysis suggested that paclitaxel is slightly more cost effective when compared with the current UK treatments than with the trial's control treatment.

**Example 7: Leukotriene Antagonists for Asthma**

The existing evidence base consisted almost entirely of trials versus placebo. The realistic treatment alternatives have not yet been able to be examined within a long term trial. This example<sup>[24]</sup> is discussed further in the sections on sensitivity analysis (section 5.5) and pre-trial modelling (section 5.4).

Buxton et al.<sup>[12]</sup> categorise these role of combining data sources as ‘synthesising results where no relevant clinical trial exists’ and distinguish a further form, ‘broader studies to test the impact of many variables’.

**5.2.3 Modelling Broader Systems**

The possible consequences of a successful intervention may be quite diffuse. A successful programme for reducing smoking, for example, has consequences for many areas of medicine and surgery and it is inconceivable that a single trial could encompass all of these. Broader studies can model systems by including the effectiveness of individual treatments on the clinical pathway, the adverse effects, the different utilities at different points in the system and the cost of the pathways.

**Example 8: Treatments to Reduce Obesity**

The long term impacts of obesity are reasonably well described from the evidence base. Increased risks of coronary heart disease, diabetes mellitus, etc. are well established. SCHARR is currently modelling the potential long term impact of therapies to reduce obesity. This is a useful and indeed necessary approach given the alternative of very large and long term observational studies.

**5.3 Generalising Results from One Specific Context to Others**

Perhaps the most important reason for modelling is the multitude of different settings in which a particular technology can be applied. Two aspects of generalising between contexts are considered here. The first is from trials into practice, and the second is from one place to another.

**5.3.1 Trials Into Practice**

Results from highly controlled trials with selected patients and a strict protocol may have only partial relevance to a Health Authority wishing to make a decision to purchase the technology in question for everyday use.

**Example 9: Cost Effectiveness of Selective Serotonin Reuptake Inhibitors (SSRIs) for the Treatment of Depression**

There are hundreds, perhaps thousands, of trials that have reported on the efficacy of drugs for the treatment of depression. However, a broader systems modelling of the treatment of depression at a community level<sup>[25]</sup> demonstrates that 4 key variables for effectiveness are: (i) the GP’s ability to recognise depression; (ii) the appropriate treatment being prescribed (subtherapeutic dosage); (iii) patient compliance; and (iv) continuation with the prescribed treatment. Effectiveness and cost effectiveness in practice can be substantially different from that reported in short term RCTs.

Drugs that are similar in efficacy are potentially very different in their effectiveness if they have different effects on patient and doctor behaviour. Modelling focuses us on obtaining more from the trials carried out by examining the adverse effect profiles of drugs and their behavioural effects and allows us to explore the likely consequences.

**5.3.2 Generalising Between Locations**

Even a passing acquaintance with health service delivery patterns demonstrates large variations in practice both between and within countries. Variations can be seen in the absolute level of resources available, for example capital equipment and staff-mix including, in the extreme, certain categories of staff completely absent from some places. Secondly, there can be considerable variations in patient management and, thirdly, there are variations in costs and prices. Modelling allows the tailoring of a robust set of clinical results to the different modes of organisation in particular locations.

**Example 10: Acute Rejection in Renal Transplantation**

US protocols for renal transplantation target an inpatient length of stay of 8 days, whereas the

German target is 40 days. A drug to reduce rejections has very different resource consequences and its cost effectiveness will vary significantly between countries.<sup>[26]</sup> The use of US cost-effectiveness results by a decision-maker in Germany would be wholly inappropriate and a model is a necessity.

#### Example 11: Cost Effectiveness of Statins in the UK

Generalisation between locations was also important when analysing the potential impact of statin therapies in the UK.<sup>[27,28]</sup> The Scandinavian trial used to estimate benefits reported reductions in the use of coronary artery bypass grafts, etc. However, the UK setting has an intervention rate of around 50% of the Scandinavian rate. An adjustment for the percentage reduction in open-heart surgery interventions was made using estimates of UK intervention rates. While this adjustment reduced the costs avoided, the analysis still showed statins as cost effective for the risk groups defined.

### 5.4 Modelling Prior to Trials and Studies

The value of modelling prior to clinical trials and studies lies in informing study design and in setting priorities for future research. It can be used to:

- help to generate hypotheses that can be tested by trials
- decide on the key outcome variables to be measured
- quantify the potential value of trials.

Such approaches are useful both for government agencies such as the NHS Health Technology Assessment (HTA) Programme and for pharmaceutical companies who need to prioritise further research and development work.

We are currently undertaking a formal systematic review for the HTA on 'the use of modelling and planning and prioritising clinical trials'.<sup>[29]</sup> There is much discussion of methodologies in the literature, but fewer concrete examples exist. For example, at the midway point in the review we had identified around 30 papers that discuss the role of modelling in prioritisation of research but fewer than 10 studies that explicitly claim to do the task for real. The methodology papers have several varia-

tions on a process for using modelling to inform research priorities:

- develop and use a model to estimate costs and outcomes in the setting of interest
- examine the sensitivity of outcomes and costs to the uncertainty in input values
- identify key uncertain variables and relationships for research
- prioritise research by comparing the costs of obtaining information on the real parameter value with the potential service cost and health consequences of the existing strategy.

Buxton et al.<sup>[12]</sup> recommend modelling as the 'tool of first resort' for early economic evaluation, both for pharmaceutical companies and government agencies. Sculpher et al.<sup>[30]</sup> reviewed the iterative relationship between modelling work and the process of health technology assessment using a 4-stage analysis. At stage I, 'early developmental work on a technology', systematic review and informed clinical opinion are proposed as the key tools in economic evaluation. While this is true, modelling can be used even here to inform the design of the next study or prioritise alternative interventions for research.

At stage II, 'maturing innovation', decision analytic modelling techniques are used to provide a coherent framework and means for synthesising data from various sources. These can be inexpensive, updateable and useful for sensitivity and threshold analysis, e.g. what level of incidence for a particular disease is likely to make the intervention worth while.

In stage III, 'close to widespread diffusion', modelling can inform data collection within RCTs. For example, if modelling indicates that the cost effectiveness of a technology is sensitive to 1 or 2 parameters, then stage III RCTs can focus on these variables and the model can be updated when the trial results are available.

At stage IV, 'moving into practice', modelling is mostly concerned with the extrapolation of the results of earlier analyses, incorporating clinical, epidemiological and economic data. Depending on the cost and the remaining clinical uncertainty after

this modelling work it may still be necessary to undertake longer term RCTs to provide more reliable estimates of the overall gains in survival. In the case of the modelling of cholesterol-lowering drugs, such trials were considered necessary and have now begun to be reported with their results being incorporated into further models.

#### Example 12: Quantifying the Value of Research in Reducing Uncertainty

A recent paper by Fenwick et al.<sup>[31]</sup> builds on the theoretical work of Claxton and Posnett,<sup>[32]</sup> Detsky<sup>[33]</sup> and others. The example is a relatively simple disease with 2 alternative forms of diagnostic test. The modellers used literature and clinical opinion to populate a decision analytic model with central estimates and ranges of uncertainty for the key parameters. Three approaches are described:

- deterministic analysis with 1-way sensitivity – yielding base case cost-effectiveness estimates and the degree of uncertainty in the results
- stochastic sensitivity analysis – using estimated distributions for each parameter and Monte Carlo simulation techniques to estimate the full uncertainty in expected costs and effects
- value of information analysis – calculating the expected value of perfect information for selected parameters or combinations of them within the model.

The results showed that further research on unit costs had little value, research on the accuracy of the alternative tests had some value, but that the most valuable research would inform the relative utility of the various disease state consequences. To address this key uncertainty did not require an RCT design.

#### Example 13: The Economic Value of a Pragmatic Trial

In a modelling study for a pharmaceutical company we undertook a detailed sensitivity analysis of key parameters in a large clinical pathways model. This examined prevalence, different treatment pathways, uptake (in practice as opposed to within trials), the potential impact of adverse effects, uncertainty in the outcomes of different interventions, etc. The results of the multiway

Monte Carlo sensitivity analyses suggested that there was a low probability that a large pragmatic trial would show cost effectiveness for the pharmaceutical company's product. This, together with the large costs of the intended pragmatic trial and its potential replication by government-funded research, helped the company to decide to avoid over £1 million of pragmatic trial investment.

#### Example 14: Early Decisions on Research and Development Investment for Pharmaceuticals

A pharmaceutical company required an examination of the relative priorities of research and development for pharmaceutical interventions at various different stages of a chronic disease. The company was interested in assessing the potential value of new products both to the UK NHS and to itself. Possibilities included interventions early in the asymptomatic phase of the disease or later to reduce deterioration in the symptomatic phase, interventions to improve the efficacy of surgery or interventions even later in the management of severe disease. A clinical pathways model was developed from prevalence through screening and diagnosis, to early, medium, surgical and late-stage interventions. The model assessed the impact of postulated new interventions on flows down particular clinical pathways and on NHS costs. Threshold analysis gave an indication of the costs and the efficacy required of a proposed intervention in order to improve on existing first-line treatments.

### 5.5 Sensitivity Analysis and Modelling Uncertainty

A recent review by Briggs et al.<sup>[34]</sup> identified 4 types of uncertainty and corresponds closely with the roles identified for modelling:

- uncertainty in the sample data – either resource use data or effectiveness of treatments
- the generalisability of results – trials with atypical management or different settings
- extrapolation – to the longer term or intermediate clinical end-points to final outcomes
- uncertainty relating to analytical methods – including, for example: (i) incorporating time preference; (ii) methods of valuation of the consequences of interventions; (iii) instruments to

value nonresource consequences; and (iv) whether to include indirect costs and costs of healthcare due to unrelated illness, during extra years of life.

Modelling is an explicit methodology for exploring uncertainty in each case. Conclusions are clearly problematic where source data are inaccurate. However, in essence, this is not a modelling problem but a data problem that any other method would share. Eddy<sup>[3]</sup> argues that the criticisms of modelling when there are poor data are ‘testimony to the fact that modelling has made this, previously obscured, fact clear’. Modelling allows us to test sensitivity to indicate the importance of poor source data and hence to estimate the value of further data collection.

The Briggs et al.<sup>[34]</sup> review identifies 4 approaches to sensitivity analysis; these are discussed in the next 4 sections.

#### **5.5.1 Type 1: Simple Sensitivity Analysis**

This assesses the impact of changing one variable.

Example 15: High Dose Chemotherapy and Stem Cell Transplantation

The effectiveness of treatments of Hodgkin’s disease, non-Hodgkin’s lymphoma and multiple myeloma required both extrapolation of trial evidence to the longer time period and sensitivity analysis. Each of these diseases had a single RCT as the evidence base. The trial evidence suggested that there was additional survival benefit and that in the case of Hodgkins disease and non-Hodgkins lymphoma that this might reach a plateau. A sensitivity analysis of a 5-, 10- and 20-year continuation of the survival benefit allowed an analysis of cost effectiveness. The conclusion was that high dose chemotherapy provides a cost-effectiveness ratio of between £12 000 and £18 000 per life year gained, which almost halves if 10-year survival estimates are assumed.<sup>[35,36]</sup>

#### **5.5.2 Type 2: Threshold Analysis**

This is calculating the value a variable would need to reach in order to change the cost-effectiveness policy.

Example 7 Revisited: Leukotriene Antagonists for Asthma

The uncertainty analysis for leukotrienes calculated the threshold treatment effect required to prove leukotriene cost effectiveness. The uncertainty suggested that the intervention was unproven at this stage.<sup>[24]</sup> There has been a recent call by the NHS HTA to undertake a clinical trial of leukotrienes in practice. This early economic evaluation and threshold analysis will inform our bid for the design of the clinical trial.

#### **5.5.3 Type 3: Extremes Analysis**

This is assessing the impact of moving one or more variables to its potential extreme value.

#### **5.5.4 Type 4: Probabilistic Monte Carlo Sensitivity Analysis**

This examines the effect of multiway variation in the input parameters in a model.

Example 16: Modelling the Routing of Severe Head Injuries

Our recent modelling study analysed whether patients with serious head injuries should be routed direct to a local hospital without neurosurgical facilities or to more distant neurosurgical centres.<sup>[37]</sup> A mathematical model of the case-mix, geography and survival parameters for a specific centre was developed that combined audit databases, published literature and clinical opinion. Due to the large number of clinical estimates that were necessary, the modelling made significant use of sensitivity analysis. One-way sensitivity identified the key uncertain variables affecting the policy decision. Monte Carlo multiway analysis allowed assessment of 10 alternative policies, of which 4 performed consistently well. The exercise also highlighted significant gaps in the knowledge concerning some variables and quantified how sensitive the model output was to these variables. The results could be used to inform sample size calculations for future observational studies.

#### **5.5.5 Type 5: Structural Sensitivity Analysis**

To the preceding 4 types of sensitivity analysis should be added structural sensitivity analysis – such as assuming different functional forms for extrapolating outcomes, e.g. constant benefits, linear extrapolation, time-dependent decay functions.

Structural sensitivity analysis is an area that is relatively neglected and we would argue that it should be routinely considered alongside the others.

## 5.6 Further Aspects of the Value of Modelling

### 5.6.1 Value as Communication Tools

Models have 3 principal advantages as communication tools. First, models are explicit – precise definitions, assumptions and estimates that are open to view and criticism. Secondly, they provide a framework for formation of a consensus. The Trent DEC models have either provided a framework for the formation of a consensus or have been useful to identify the differences of opinion explicitly. Models can be used to focus a group's energy to produce agreement on issues such as the options to be assessed, the definitions and the structure of the problem, the basic factual evidence, the value and uncertainty of parameters. Finally models can also be dissemination tools. With policy made, models can be disseminated for local users, either to refine the analysis and revisit the policy with local data or to be part of a programme of education and influence to improve implementation.

#### Example 11 Revisited: Disseminating Statins Analysis

The statins cost-effectiveness model is available on the World Wide Web.<sup>[38]</sup> This has allowed the dissemination of the analysis to any Health Authority in the UK and the implementation of the analysis with local population and coronary heart disease prevalence data. This helps to inform local policy and likely need for each risk group.

### 5.6.2 Value in Problem Structuring – Conceptual Modelling and Subjective Information

Even before attempting quantification, conceptual models are useful to identify important factors or variables and define or postulate relationships between those variables, i.e. how one affects another. Such '*a priori* structuring' is often accompanied by the gathering and analysis of subjective clinical opinion. This ranges from the crude and simplistic 'give us your best guess', through to systems to assess the internal coherence of responses,

to the formalised Delphi techniques and systematic approaches that aim for a research-orientated level of rigour. Quantified models can take the results of these approaches as inputs. They are also useful to identify questions and assess the validity or uncertainty of subjective information.

### 5.6.3 Informing the Problem Situation when Hard Data are Impossible to Obtain

When hard data are impossible to obtain, modelling is used to structure the decision problem and test the policy decision's sensitivity to assumptions about what might happen. Buxton et al.<sup>[12]</sup> consider vaccination for swine influenza, a situation where it was necessary to decide on vaccination or otherwise before the scale of the potential epidemic or the efficacy of the vaccine was clear. Modelling tested the sensitivity of various assumptions and showed that vaccination was very unlikely to be cost effective. It is almost impossible to do a meaningful small scale trial of national vaccination. The efficacy of the drug, take-up rates of the vaccination and potential reductions in infection rates as a consequence of herd immunity must be modelled.

Modelling is also necessary when it is ethically or politically difficult to gather further data, e.g. the intervention is already embedded in current practice.

#### Example 17: Partial Hepatectomy for Liver Metastases

This intervention is used by many clinicians across the world with increasing sophistication. There are no published RCTs. 21 independent case series were reviewed. In many of the series the non-resected comparators were unsuitable for surgery and were certainly not an unbiased comparator group. The leading case series in terms of size, patient mix comparators and surgical techniques was used to extrapolate outcomes using scenarios of 5-, 10- and 20-year continued benefit. The conclusions suggested that the intervention was very cost effective.<sup>[39]</sup>

#### Example 18: Magnetic Resonance Imaging (MRI) in the Management of Knee Disorders

Again there were no RCTs of MRI. Some case studies suggest MRI is effective and others that it adds nothing. Scenario analysis of the case

series showed that there was more impact when general orthopaedic surgeons used MRI than with knee specialists. The analysis did not find enough evidence to conclude on cost effectiveness and recommended a local audit of MRI practice.<sup>[40]</sup>

## 6. Discussion

### 6.1 The Need for Quality Assurance and Validation

Modelling plays, and must play, a key role in economic evaluation of healthcare technologies if that evaluation is to be of value to policy-makers and decision-takers. This provides a challenge to the members of NICE in the UK, and similar bodies around the world, in deciding when a conclusion derived from a model is valid. The challenge is significant. Clinical trials have been with us sufficiently long to be well formalised and most clinical scientists can distinguish a good trial from a bad one. Modelling for economic evaluation is significantly more complex than designing trials.

Eddy<sup>[3]</sup> identifies the following key limitations in modelling. First, it does not provide new observations. If based on incorrect clinical judgements, modelling will perpetuate any of these errors (garbage in – garbage out). Models can be poorly designed (e.g. decision trees can be wrongly structured or use biased expert opinion). Oversimplification (the most common error) can occur by omitting important variables, squeezing the problem into a familiar or convenient mathematical form or assuming the outcomes assessed by the model are the only ones of interest. Finally, results can be misinterpreted and decision-makers may fail to appreciate the degree of uncertainty in the results. Sheldon<sup>[11]</sup> and Buxton et al.<sup>[12]</sup> also identify different versions of these same problems with modelling approaches and give examples where models have been shown subsequently by trials to be wrong (as of course have many trials themselves).

### 6.2 Validation

The understanding and development of methodologies for quality assurance and validation is at an

early stage in the context of health technology assessment. In the disciplines of operational research and applied mathematical modelling in general, there is a large and varied literature discussing model validation. This focuses both on the hard numbers and the softer processes of model development and problem structuring and ranges from the deeply philosophical to practical rules of thumb.<sup>[41-47]</sup>

Within HTA, Eddy<sup>[3]</sup> has provided some guidance to good practice. First, he identifies 4 different orders of validation:

(i) Expert concurrence – the model should make sense to people with knowledge of the disease (the right factors are included, the mathematical relationships are intuitive, the data sources are reasonable).

(ii) Internal validity – the model should match the results of the source data used to construct it. Failure to do so gives a strong indication that the structure is wrong, i.e. the wrong relationship assumptions have been made.

(iii) Predictions agree with nonsource data – the model is used to make predictions that are then cross-checked from nonsource data. This is clearly easier when modelling usual care since there is more data available. For a new therapy, Eddy notes that there is a trade-off between saving a partition of the data for validation purposes versus using as much of the available evidence as possible to construct the model in the first place.

(iv) Predict – experiment – compare – this highest level test of validation is a useful approach where it is possible to undertake the experiment. One problem is that the actual conditions being experimented upon can be different from those assumed within the model, but if the model is well structured then the important experimental conditions should be key parameters within the model.

An early checklist for the quality of modelling was produced by Eddy<sup>[3]</sup> (table I) and further developed by Sonnenburg.<sup>[48]</sup> An increasing element of our own work is in the critical review of models.

Examples 19 and 20: Critical Reviews

A model of vaccination for pertussis from another country was critically reviewed (Beard S,

personal communication). It was shown to have very different infection or 'attack rates' compared with the UK. Our review report suggested and subsequently implemented a revision of these rates to translate to the UK experience. We also developed the model to reflect adequately the dynamics of infection, including herd immunity (i.e. the higher the population coverage of vaccination, the lower the 'attack rate') and the influence of adult infection.

A similar model review on vaccination for hepatitis B discovered that the assumptions about the effects of herd immunity were not adequately covered by the existing model. A full scale revision of the model was required.

Assessing models will severely tax our normal peer review procedures. Constraints on length mean that all the information needed to assess the quality of a model cannot be printed in the methods section of a typical paper in a clinical or economic journal. Recent suggestions of a published short paper<sup>[49]</sup> together with a long paper or even the model itself available electronically over the World Wide Web would represent a step forward.

Our own views echo those of Rittenhouse,<sup>[5]</sup> Eddy<sup>[3,4]</sup> and others. Ultimately, good and bad analysis will show themselves via the scrutiny of qualified reviewers. Transparency on the part of analysts is a necessary, though not sufficient, condition for quality and validity. Understanding how to review a model contributes immeasurably to the goal of not being misled by them. Researchers, policy-makers and journals will need to agree on how the validation, peer review and publication processes will be handled if the degree of trust in modelling results is to increase.

## 7. Conclusions

From Markov to Monte Carlo, from cost adjustments to forecasting long term outcomes, the variety of frameworks, purposes and roles of modelling in economic evaluation is substantial. This paper, reviewing 20 examples mostly from our own policy decision support work, shows the need for modelling and its value in the systematic and transparent synthesis of evidence to support policy de-

**Table I.** Checklist for the quality of modelling produced by Eddy<sup>[3]</sup>

**Each report of a technology assessment employing a mathematical model should contain the following elements:**

- a statement of the problem
- a description of the relevant factors and outcomes
- a description of the model
- a description of data sources (including subjective estimates), with a description of the strength and weaknesses of each source
- a list of assumptions pertaining to: (i) the structure of the model (e.g. factors included, relationships, and distributions); and (ii) the data
- a list of the parameter values that will be used for a base case analysis, and a list of the ranges in those values that represent appropriate confidence limits and that will be used in a sensitivity analysis
- the results derived from applying the model for the base case
- the results of the sensitivity analyses
- a discussion of how the modelling assumptions might affect the results, indicating both the direction of the bias and the approximate magnitude of the effect
- a description of the validation method and results
- a description of the settings to which the results of the analysis can be applied and a list of main factors that could limit the applicability of the results
- a description of research in progress that could yield new data that could alter the results of the analysis

**If the analysis recommends a policy, the report should also contain the following elements:**

- a list of the outcomes that required value judgements
  - a description of the values assessed for those outcomes
  - a description of the sources of those values
- the policy recommendation
- a description of the sensitivity of the recommendation to variations in the values
  - a description of the settings to which the recommendations apply

isions. The challenge for reimbursement authorities is how to decide when a conclusion derived from a model is valid. The challenge to the academic community, to provide a consensus on what represents good quality modelling, has never been more urgent. It is to be hoped that the considerations of this conference will represent steps towards such a consensus.

## Acknowledgements

This paper could not have been written without the Trent Institute Acute Purchasing Working Group and the staff of the Operational Research Section of SchARR. Thanks are particularly due to Chris McCabe for initiating the consensus

conference and Steve Gallivan who was discussant for the paper.

## References

- Canadian Co-ordinating Office for Health Technology Assessment. Guidelines for the economic evaluation of pharmaceuticals. 1st ed. Ottawa: CCOHTA, 1994
- Department of Health, Housing and Community Services. Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Australian Government Publishing Service, 1995
- Eddy DM. Assessing medical technology. In: Eddy DM, editor. Technology assessment: the role of mathematical modelling. Washington, DC: National Academy Press, 1985: 144-75
- Eddy D. Should we change the rules for evaluating medical technologies? In: Gelijns A, editor. Modern methods of clinical investigation. Washington, DC: National Academy Press, 1990
- Rittenhouse B. Uses of models in economic evaluations of medicines and other health technologies. London: Office of Health Economics, 1996
- Luce BR. Policy implications of modelling the cost-effectiveness of health care technologies. *Drug Info J* 1995; 29: 1469-75
- Kassirer JP, Angell M. The journal's policy on cost-effectiveness analyses [editorial]. *N Engl J Med* 1994; 331 (10): 669-70
- Task Force on Principles for Economic Analysis of Health Care Technology. Economic analysis of health care technology: a report on principles. *Ann Intern Med* 1995; 122: 60-9
- Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world: effectiveness versus efficacy studies. *Pharmacoeconomics* 1999; 5 (5): 423-34
- Rizzo JD, Powe NR. Methodological hurdles in conducting pharmacoeconomic analyses. *Pharmacoeconomics* 1999; 15 (4): 339-55
- Sheldon TA. Problems of using modelling in the economic evaluation of health care. *Health Econ* 1996; 5: 1-11
- Buxton MJ, Drummond MF, Van Hout BA, et al. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997; 6: 217-27
- OHE Briefing. The pros and cons of modelling in economic evaluation. London: Office of Health Economics, 1997
- Office of Health Economics. Handbook of the Office of Health Economics database of health economic evaluations. London: Office of Health Economics, 1998
- McGuire WP, Hoskins WJ, Brady MF, et al. Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *N Engl J Med* 1996; 334 (1): 1-6
- Beard SM, Coleman R, Radford J, et al. The use of cisplatin and paclitaxel as a first line treatment in ovarian cancer. Sheffield: Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 1997
- Chilcott J, Golightly P, Jefferson D, et al. The use of riluzole in the treatment of amyotrophic lateral sclerosis. Sheffield: Trent Institute for Health Services Research, 1997
- Cummings C. The use of olanzapine as a first and second choice treatment in schizophrenia. Birmingham: Department of Public Health and Epidemiology, University of Birmingham, 1998
- Beard SM, Brewin J, Packham C, et al. A review of the use of current atypical antipsychotics in the treatment of schizophrenia. Sheffield: Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 1998
- US General Accounting Office. Cross-design synthesis: a new strategy for medical effectiveness research [US GAO/PEMD-92-18]. Washington, DC: US General Accounting Office, 1992
- Drummond MF, Davies L. Economic analysis alongside clinical trials. *Int J Technol Assess Health Care* 1991; 7: 561-73
- Knight C. Modelling *Helicobacter pylori* eradication: the HEALS model. Project report to Abbott Pharmaceutical. Sheffield: School of Health and Related Research, University of Sheffield, 1999
- European Society of Primary Care Gastroenterology (ESPCG). Pan-European consensus on patient management strategies for *H. pylori*: the management of *H. pylori* infection [bulletin]. ESPCG Pan-European Consensus on Patient Management Strategies for *H. pylori*; 1998 May 9-10; Zurich
- Stevenson MD, Richards RG, Beard SM. The health economics of asthma and the role of antileukotrienes in the treatment of chronic asthma. Sheffield: Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 1999
- Beard SM, Ward SE, Brennan A. SchHARR Project Report to SmithKline Beecham: sensitivity analysis of the 'ADEPT' Depression Management Model. Sheffield: School of Health and Related Research, University of Sheffield, 1997
- Akehurst RL, Chilcott, JB, Holmes MW, et al. The economic implications of the use of basiliximab versus placebo for the control of acute cellular rejection in renal allograft recipients. Sheffield: School of Health and Related Research, University of Sheffield, 1999
- Pickin MD, Payne JN, Haq IU, et al. Statin therapy/HMG-CoA reductase inhibitor treatment in the prevention of coronary heart disease. Sheffield: Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 1996
- University of York. Cholesterol and coronary heart disease: screening and treatment. *Eff Health Care* 1998; 4 (11): 1-16
- Chilcott JB, Brennan A. Systematic review of the use of modelling in planning and prioritising clinical trials. International Society for Technology Assessment in Health Care (ISTAHC) Conference; 1999 Jun 20-23; Edinburgh
- Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. *J Health Serv Res Policy* 1997; 2: 26-30
- Fenwick E, Sculpher M, Claxton K, et al. The role of decision analytical modelling improving the efficiency and relevance of health technology assessment. York: Health Economics Study Group, 1998
- Claxton K, Posnett J. An economic approach to clinical trial design and research priority setting. *Health Econ* 1996; 5 (6): 513-24
- Detsky AS. Using cost-effectiveness analysis to improve the efficiency of allocating funds to clinical trials. *Stat Med* 1990; 9: 173-84
- Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Econ* 1994; 3: 95-104
- Beard SM, Lorigan P, Sampson F, et al. The effectiveness of high dose chemotherapy with autologous stem cell transplantation in the treatment of Hodgkin's disease and non-Hodgkin's lymphoma. Sheffield: Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 1998.
- Beard SM, Sampson FC, Vandenberghe E, et al. The effectiveness of high dose chemotherapy with autologous stem cell/bone

- marrow transplantation in the treatment of multiple myeloma. Sheffield: Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 1998
37. Stevenson MD, Beard SM, Oakley PA, et al. Modelling of the potential impact of a regional trauma system in patients with severe head injury. Sheffield: School of Health and Related Research, University of Sheffield, 1998
  38. The SchARR statins prescribing model. Available from URL: <http://www.shef.ac.uk/~scharr/or/statins/> [Accessed 2000 Mar 20]
  39. Beard SM, Holmes M, Majeed A, et al. Hepatic resection as a treatment for liver metastases in colorectal cancer. Sheffield: Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 1999
  40. Beard SM, Perez I, Touch S, et al. Magnetic resonance imaging (MRI) in the management of knee disorders. Sheffield: Trent Institute for Health Services Research, Universities of Leicester, Nottingham and Sheffield, 1998
  41. Robinson S. Successful simulation: a practical approach to simulation projects. London: McGraw-Hill, 1994
  42. Robinson S, Pidd M. Provider and customer expectations of successful simulation projects. *J Oper Res Soc* 1998; 49: 200-9
  43. Gass SI. OR in the real world: how things go wrong. *Comput Oper Res* 1991; 18 (7): 629-32
  44. Fossett CA, Harrison D, Weintrob H. An assessment procedure for simulation models: a case study. *Operations Res Soc Am* 1991; 39 (5): 710-24
  45. Finlay PN, Wilson JM. Orders of validation in mathematical modelling. *J Oper Res Soc* 1990; 41 (2): 103-9
  46. Gass SI. Model accreditation: a rationale and process for determining a numerical rating. *Eur J Oper Res* 1993; 66: 250-8
  47. Finlay PN, Forsey GJ, Wilson JM. The validation of expert systems: contrasts with traditional methods. *J Oper Res Soc* 1988; 39 (10): 933-8
  48. Sonnenberg FA, Roberts MS, Tsevat J, et al. Toward a peer review process for medical decision analysis models. *Med Care* 1994; 32 (7): JS52-64
  49. Delamothe T, Smith R. Moving beyond journals: the future arrives with a crash. *BMJ* 1999; 318: 1637-9
- 
- Correspondence and offprints: Mr *Alan Brennan*, School of Health and Related Research (SchARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, England.  
E-mail: [a.brennan@sheffield.ac.uk](mailto:a.brennan@sheffield.ac.uk)