No 33 May 1997

OHE briefing

THE PROS AND CONS OF MODELLING IN ECONOMIC EVALUATION

INTRODUCTION

The use of models and modelling in the economic assessment of health care technologies including pharmaceuticals to help answer the cost-effectiveness questions (does it work, and if so does it represent value-for-money?) is controversial. We use the term model in two quite different ways (Rittenhouse, 1996). It can be any artificial simplification of reality designed to enable us to better understand the world. A road map would fall into this category as would a randomised controlled trial. It is the second meaning of the term model that is more controversial – where the simplification of reality includes the use of techniques to combine data from different sources, and, usually, the use of assumptions to enable extrapolation from the combined data or to fill gaps within the required data set. In this case an important issue is whether models are being used to predict cost-effectiveness or to highlight issues on which judgements must be made or further research commissioned.

This paper summarises the presentations and comments of the panel speakers to the issues raised by criticisms of modelling, considering:

- definitions of models;
- the use of models outside of economic evaluation;
- the uses of models in economic evaluation;
- why modelling in economic evaluations has become such an important issue;
- the benefits of models in economic evaluation;
- the problems with modelling in economic evaluation;
- the alternatives to modelling;
- the way forward.

The paper ends with concluding comments on the issues raised by the session.

Professor Drummond introduced the session by highlighting the two alternative approaches to economic evaluation:

- the 'trial-based' approach with concurrent data collection (for example on resource use) alongside a clinical trial;
- the integrative study, with modelling of, or a synthesis of, data from a number of sources.

He noted that there had been significant recent criticism of integrative studies, for example by the New England Journal of Medicine (NEJM) and the US Food and Drug Administration (FDA). Their sceptical view of modelling was not shared by all health economists and decision makers. Many argue that in the absence of good trial-based information it was difficult to help decision makers without the use of models. This raises the question as to whether modelling is a stop-gap while we wait for clinical trial-based information, or an inevitable, and perhaps desirable, part of economic evaluation? A separate question is

This OHE Briefing summarises the presentations and discussion at the session on 'Modelling in Economic Evaluation' at the Conference of the International Society for Technology Assessment in Health Care (ISTAHC) in San Francisco on June 26 1996.

The participants were:

Professor Bernard Bloom, *University of Pennsylvania, USA*
Professor Martin Buxton, *Director of the Health Economics Research Group, University of Brunel, UK*
Professor Michael Drummond, *Director of the Centre for Health Economics, University of York, UK*
Dr Bryan Luce, *Chief Executive of MEDTAP International, USA*
Professor Trevor Sheldon, *Director of the NHS Centre for Reviews and Dissemination, University of York, UK*

whether modelling risks introducing hidden bias, and if this bias can be identified? Do we have the instruments to judge models – or must all modelling carry a health warning? The ISTAHC session was designed to explore these different perspectives on the role of modelling.

DEFINING MODELS

Professor Sheldon described modelling as a way of representing the complexity of the real world in a more simple and comprehensible form. It involved a theoretical description that helps us to understand how a system or a process works or might work, simplifying reality in order to try and understand that reality a bit more. Professor Bloom focused on the modelling technique known as decision analysis. He defined this as a tool for expressing the known, observed or expected reality in mathematical terms. In the context of health care this could be a disease state, a treatment or an entire episode of care. It allows us to simulate or estimate various realities in order to help predict the future. Clinical and economic features can be defined based on known and estimated inputs and interactions. Of course, the accuracy of the prediction can only be proven with time and will be dependent on the assumptions and inputs used as the basis for the probabilities and costs of events occurring. Decision analysis, like every research tool, works best when the assumptions, logic and inputs are grounded in fact, but often the greatest value of decision analysis is when there are too few known inputs to make an easy decision.

> *'Decision analysis, like every research tool, works best when the assumptions, logic and inputs are grounded in fact, but often the greatest value of decision analysis is when there are too few known inputs to make an easy decision.'*
>
> PROFESSOR BLOOM

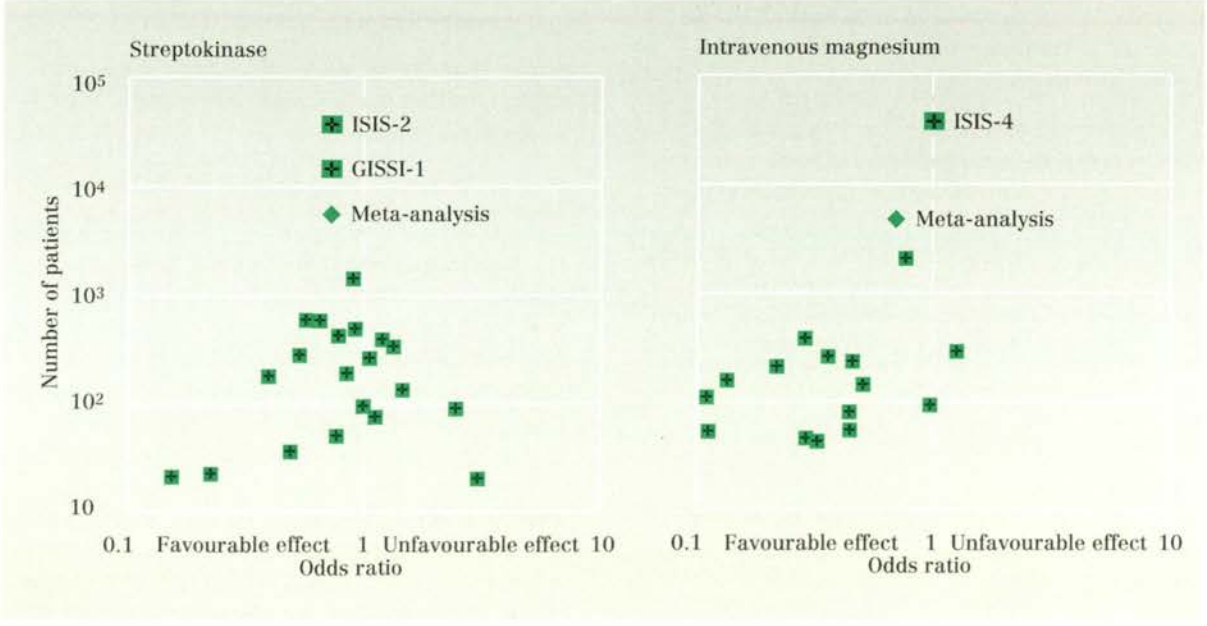
THE USE OF MODELS OUTSIDE OF ECONOMIC EVALUATION

Professor Sheldon commented that forms of modelling were often used in health services research outside of economic evaluation. For example: it occurs in data collection (we simplify whenever we code data with some sort of categorical variable to slot people into pigeon holes); when using regression analysis; and when calculating an odds ratio.

He gave three more substantive examples of the use of models in clinical research outside of economic evaluation, introducing a note of caution as to their use:

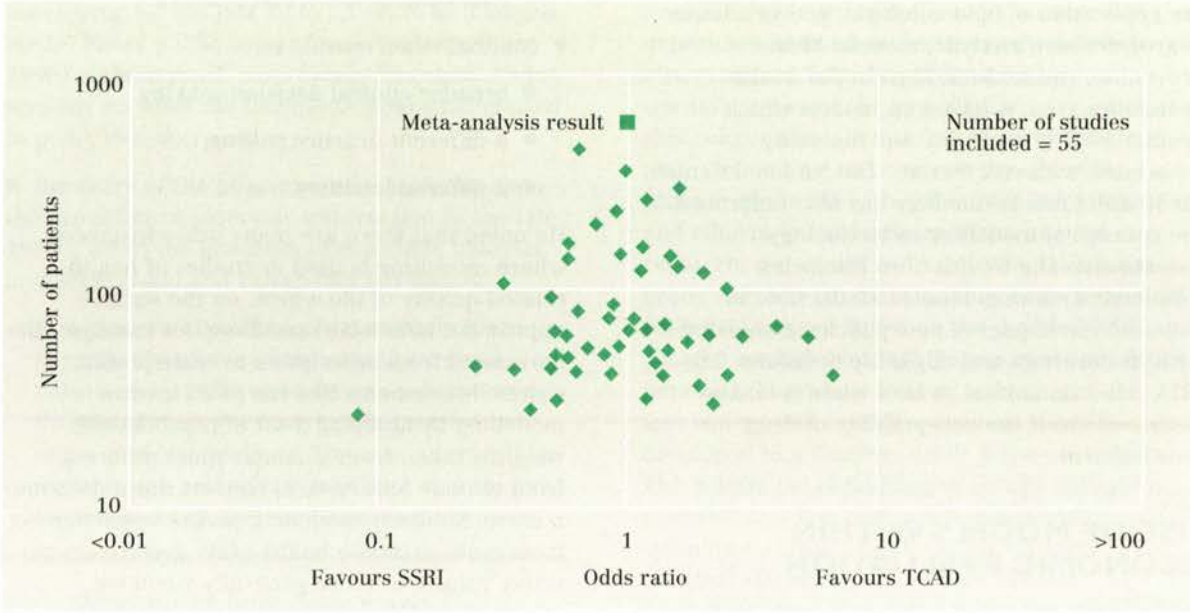
- Subgroup analysis – when we attempt to model whether an intervention is of use to certain types of patients within a larger study. The DICE study, carried out in Oxford, showed that you can identify lots of apparently 'significant' subgroups that are the result of chance (Counsell et al, 1994);

Figure 1



Source: Derived from Egger and Davey-Smith, 1995.

Figure 2 **Funnel plot of SSRI vs TCAD total drop out rates**



Source: Derived from Song et al, 1993.

- Meta-analysis – a form of modelling that involves combining evidence from different studies to obtain more precise or unbiased estimates of effect. There can be problems – particularly with publication bias (small studies are more likely to be completed, submitted and accepted for publication if they show significant positive effects). If, for example, we consider the use of streptokinase at the time of a heart attack, we can plot the results of existing trials. These included small trials, alongside very large randomised controlled trials (see Figure 1). We can also plot the overall, meta-analysis, estimate. The result is called a 'funnel plot', and in this case it is quite symmetrical. The overall estimate is not the result of a preponderance of small trials which would most probably reflect publication bias. On the other hand, Figure 1 also shows the meta-analysis for intravenous magnesium. Here there is a preponderance of small trials showing a lot of benefit, indicating likely publication bias (Egger and Davey-Smith, 1995). When a large trial was conducted it showed no effect (ISIS-4, 1995). In this case the meta-analysis produced an incorrect estimate because the modelling technique had poor data to use.

- We have seen similar problems in more commercially important areas like that of selective seratonin re-uptake inhibitor (SSRI) anti depressants versus tricyclics (TCAD). The result of the meta-analysis by Song et al, (1993), which examined drop out rates, is given in Figure 2. There is symmetry with an overall estimate of little difference in drop out. Another meta-analysis by Montgomery et al (1994) shown in Figure 3 indicates a difference. However, it has a preponderance of small trials on one side indicating that there was publication bias. There were other trials that they did not include. Meta-analysis is a form of modelling where we are beginning to develop the techniques to check whether data and results are reliable.

- Observational data needs to be adjusted, for example, for confounding by case mix or indication. Again it is difficult to know whether we have adequately modelled to remove confounding. There are several examples where the use of observational data often combined with trials or other sorts of data have given, in retrospect, biased answers. One well known example is by Eddy et al (1988) which used complex confidence profile Bayesian techniques to combine poor data. It produced a misleading result which suggested that there was value in providing breast cancer screening to women under 50.

Dr Luce observed that use of models was widespread outside health care services research, including national economic forecasting, estimating company returns on investment, pilot and astronaut training, architectural design, and industrial engineering. In health care as well, he noted, there is a history of the use of well tested models. US based organisations like Blue Cross, Blue Shield, and the Office of Technology Assessment have used models to a great extent. The National Institutes for Health fund modelling in some of the basic work they do. The Centres for Disease Control

and Prevention have a prevention effectiveness research programme that is based primarily on the application of epidemiologic, and to a lesser degree decision analytic, models. Most prevention research, at least in the health economics area, is based on models which predict events, morbidity and mortality associated with risk factors. The National Centre for Health Care Technology has also embraced the concept of modelling in technology assessment. The Health Care Financing Administration uses models all the time to estimate the impact of new policies associated with its coverage and eligibility decisions. The FDA relies on animal models when it makes decisions about the acceptability of drug development.

USE OF MODELS WITHIN ECONOMIC EVALUATION

Professor Buxton outlined the situations where models are used in economic evaluation. There are elements of modelling in exploring:

- clinical outcomes, when moving from:
  - intermediate to final outcomes;
  - short term to long term outcomes;
  - efficacy to effectiveness;
- resources, when moving from:
  - patient management patterns to resource use;
  - resource use to cost;
  - trial context to normal practice.
- context, when moving to:
  - broader clinical decision-making;
  - a different practice setting;
  - a different country.

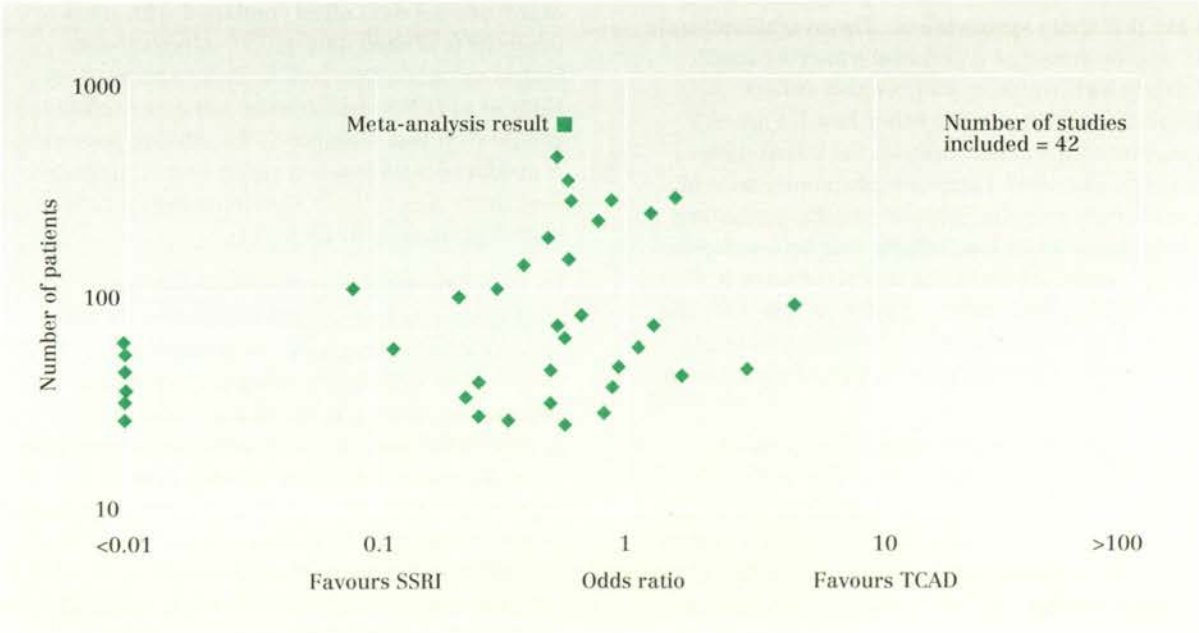
He noted that there are many other instances where modelling is used in studies of health related quality of life which, on the surface appear not to involve modelling, for example, the movement from descriptors to scale/profile scores. Instruments like the SF36 involve modelling by applying a set of psychometric weights, taken from a sample quite different from the one observed, to convert those data into a score. Similarly modelling occurs when moving from multi-attribute health state descriptors to utility values, or more generally when we construct QALYs using data from outside the patient group within the trial.

WHY MODELLING IN ECONOMIC EVALUATIONS HAS BECOME SUCH AN IMPORTANT ISSUE

Dr Luce set out three reasons:

- the stakes have been raised. The results of models are now being used as the basis of major policy decisions about pricing and reimbursement of drugs, coverage, and, initially, for rationing in

Figure 3 **Funnel plot SSRI vs TCAD drop out rates due to side effects**



Source: Derived from Montgomery et al, 1994.

the Oregon experiment (Tengs 1996). Suddenly the field of economic evaluation that had been developing for the last 20 to 25 years as a largely model based public sector funded activity (Luce, 1995) without much controversy has come under scrutiny because the technique is now being used in policy decision making;

- the entry of the pharmaceutical industry into the provision of economic information in the late 1980's led to the inclusion of economic analysis in clinical trials and raised two key issues:
  - a concern about bias. Immediately there began a discussion of the unstandardised nature of cost effectiveness analysis including the use of models which has in turn led to many sets of guidelines. There is concern that assumptions and selection of data sources can be manipulated, (efficacy data, resource use, costs/prices, discount rates). There is an opportunity for bias and with medical products industries, in particular the pharmaceutical industry, there are powerful financial incentives to encourage bias (Evans, 1995). Hence, there have appeared explicit and implicit journal editorial policies on the acceptance of this. This point was reiterated by Professor Sheldon who noted the concern as to whether modelling is sufficiently methodologically mature to ensure that the results are reliable reflections of reality, or that we can accurately assess reliability and bias;

> *'The stakes have been raised. The results of models are now being used as the basis of major policy decisions about pricing and reimbursement of drugs, coverage, and, initially, for rationing in the Oregon experiment.'*
>
> DR LUCE

  - the role of the FDA, which has a mandate in the US to control advertising and promotional claims associated with pharmaceuticals. When a model or any kind of economic analysis is used to support claims about a product's cost effectiveness, the FDA feels that it has a legal mandate to ensure that the information is not false and misleading and is within the approved indications for the medicine's use.

- There has been a major clash of cultures. Cost-effectiveness analysis comes from economic and Bayesian disciplines. These are different traditions from those followed by most clinical researchers and epidemiologists. He emphasized the point made by Professor Buxton that clinical science relies on controlled experimental data, making direct comparisons. Most economic analysis, however, relies heavily on non-experimental (time series or cross sectional) observational data. Health economics lies uneasily between the two. Dr Luce argued that this culture clash lies at the heart of debates about modelling and controversy about the attitude of the FDA, the bio-statistical community and clinical journal editors. Experimental research, especially in drug development, has brought in biostatisticians. They have a classical statistical culture based on the double-blind placebo control trial with 95 per cent confidence intervals. The whole field of technology assessment and cost effectiveness analysis developed in a fundamentally Bayesian culture. This allows the combining of results with prior probabilities. It is oriented to magnitude estimation rather than hypothesis testing. The classical statistical culture has recently dominated the debate, dismissing the contribution of Bayesian-based decision analysis.

Professor Drummond noted decision maker scepticism about the role of modelling, although many recognised that there was a role for it. Greatest concern had been expressed in the NEJM Editorial (Kassirer and Angell, 1994) which said that 'bias can compromise even original scientific studies, [by which they meant clinical trial-based studies], but we believe that the opportunities for introducing bias into economic studies are far greater given the discretionary nature of model building and data selection in these analyses'.

He noted that the FDA in its draft principles for the review of pharmacoeconomic promotion, (FDA, 1995) had argued that:

- research to substantiate pharmaco-economic claims (cost-effectiveness and quality of life claims) must meet traditional standards for adequate and well controlled studies;
- models to provide estimates of pharmaco-economic parameters should only be used when it is impractical or impossible to gather data using adequate and well-controlled studies.

Dr Luce added that the FDA may therefore be ready to recognise that models play a role, albeit a limited one in the substantiation of pharmaco-economic claims. The FDA states that 'assumptions used for model construction and in its application to the clinical setting should be explicit and must be appropriately based on rigorous scientific methods', but that 'models that seek to estimate drug effects on clinical conditions that have not been demonstrated by adequate well controlled studies or that have not been included in product labelling are unacceptable'.

Professor Drummond noted that the Australian Government in its revised cost-effectiveness guidelines (Commonwealth of Australia, 1995) proposed a two step process involving:

- primary cost-effectiveness assessment based on data from the most relevant clinical trials. 'Head-to-head' studies of the product with the comparator it would replace in clinical practice are preferred but not essential;
- any subsequent extrapolations to transfer or generalise from the trial-based evidence should be transparent.

Its logic for this approach was that trial-based comparison was the most internally valid and paid due heed to biostatistical and epidemiological rules. Hence the primary analysis should use trial-based data. However, a pragmatic approach is often required when making a decision, because trial evidence is inadequate. Hence the two stage approach was better than the two extremes of either accepting only evidence from trials or of accepting all models submitted.

Dr Luce also discussed the relevant section of the Australian guidelines. He pointed out that they acknowledge that 'frequently the randomised trials will provide insufficient information which to base a judgement about the full clinical and economic performance of the proposed drug. In these circumstances, which are a matter of judgement', a modelled economic evaluation will be useful to the Pharmaceutical Benefit Advisory Committee. Appendix J (of the Guidelines) contains advice in the circumstances where a modelled economic evaluation is likely to be informative. The Appendix states that 'modelling may be needed to address limitations in the preliminary economic evaluation based on the evidence from the randomised trials presented earlier. The list of uses of models is intended to help a sponsor decide whether a model is needed in the context of each submission. Uses include: to link surrogate outcomes; to extrapolate outcomes; to examine differences between subjects enrolled in trials and patients likely to obtain drug; to modify resource use patterns to include any relevant differences in resources not measured; to exclude protocol derived resources'.

Dr Luce added that the Canadian guidelines were more positive about modelling. They state that 'Ideally a pharmaco-economics study should report on drug effectiveness rather than efficacy. Because effectiveness data are generally not available, appropriate modelling techniques based on sound pharmaco-epidemiology are permissible. All assumptions used in such extrapolation techniques must be stated explicitly and thoroughly tested with sensitivity analysis' (CCOHTA, 1994).

THE BENEFITS OF MODELS IN ECONOMIC EVALUATION

Professor Bloom commented that information provides the underpinning for decisions – by reducing uncertainty about resource use and outcome. We need pretty good information to inform choices. In medical care, however, we often lack good prospectively derived data on the clinical, economic and quality of life processes and outcomes which can improve rational choice by enabling us to estimate with any degree of certainty the likely health and other consequences of the resource input. Decisions need to be made with insufficient information. In these circumstances we ask the question, 'what if...?' This is the case for using decision analytic modelling.

> *'We seek both scientific rigour and policy relevance. There is no point in having a very precise answer to the wrong question which is what we frequently get with randomised controlled trials... timely approximation is probably better than the ultimate answer.'*
>
> PROFESSOR BUXTON

He argued, however, that the simplified decision analytic models used in a complex health services research and decision making world are inadequate for many of the most important decisions to be made at system or national level. We tend, for rational and reasonable reasons, to simplify complex inputs and their even more complex interactions in order to understand the complex processes and outputs. We assume linearity when non-linearity in life is the norm. Models usually consist of a reduced number of variables and input/output relationships simplified to a convenient mathematical form. We use these simplifying models because we think there is no other option. This is not the case. Other fields like physics, astronomy, economics and chemistry have for years constructed models of complex systems. Medicine has begun to use such models, for example, in research on the human brain.

Professor Buxton observed that it is not often that an economist can draw on Richard Peto for support. Peto (1993) commented that 'even if you take all the best clinical trial knowledge available it still answers only a limited number of questions.'

Now we can either say 'we don't know' or 'what is the best way of bringing together what little

we know to best answer the necessary questions'. When we get to the issue of economics again the evidential gap is even greater. 'Of about 50,000 randomised trials undertaken over a 22 year period, only 121 included economic analyses.' Adams et al, (1992).

Therefore, he argued, there is a balancing act:

- we seek both scientific rigour and policy relevance. There is no point in having a precise answer to the wrong question. That is what randomised controlled trials frequently deliver. They tell us in a very precise situation with great internal validity the answer to a question that is not the real life one we want answered;

> *'It's no use simply relying on what we have observed if the decisions we have to make must take account of reasonable assumptions about what happens beyond them... Economic modelling makes assumptions explicit. If we just present the hard science individual users make their own implicit assumptions, without any ability for us to see what they are.'*
>
> PROFESSOR BUXTON

- timely approximation is probably better than the ultimate answer. We should check afterwards whether we can appraise models 20 years after the event. But waiting 20 years before acting is not very useful. We should use that evidence of what happens, subsequently, to help improve models.

It seemed to him that one of the advantages of economic modelling is that we should be able to see in an explicit way what assumptions are being made. If we don't model and say 'don't let us dirty ourselves with this soft science' then what happens is that we present only the hard science. The individual users then make their own assumptions, implicitly, without any ability for others to see what those assumptions are and to challenge them.

He referred to an example studied by Schulman et al (1991) looking at the cost-effectiveness of low-dose zidovudine therapy for asymptomatic patients with HIV infection. It did not use blood cell counts but survival at one year. They considered two models consistent with the evidence, a one time effect model, and a continuous treatment model. The first showed gain of 0.30 life years gained, the second a gain of 3.29 life years.

It could be argued that this shows that depending on the model assumptions used, you can come up with whatever result you want. It seemed to Professor Buxton, however, that this approach had the benefit of highlighting the problem for the decision maker. It is crucial to make a judgement about the long term effect. This is an example of where formal and explicit analysis is much better than implicit analysis. He again quoted Peto 'extrapolation too far may lead to the mistaken decisions about treatments, but so too may failure to extrapolate far enough' (Peto, Collins and Gray 1993). It's no use simply relying on what we have observed if the decisions we have to make must take account of reasonable assumptions about what happens beyond them.

THE PROBLEMS WITH MODELLING IN ECONOMIC EVALUATIONS

Professor Sheldon outlined problems with important uses of modelling:

- extrapolation of results to different populations, dosages and over longer time periods and from surrogate to final end points. Using models to predict the future by extrapolating beyond the time period of the trial is attractive because trials are expensive and people do want (especially in economic analysis) to say what are the benefits and costs in 10 years time. If the trial lasts for only five years what do you do? Is there a decay function and what assumptions do we make? Take the early breast cancer trialists collaboration, bringing together of all the RCTs of adjuvant systemic chemotherapy for primary breast cancer. All the experts and trialists were asked to give a prediction of what all the results are going to be if tamoxifen chemotherapy was extended for five years. Opinions ranged from a 20 per cent increase in

> *'There is no developed system of model validation or review and thus it is very hard for the user to discriminate.'*
>
> PROFESSOR SHELDON

the odds of death to a 25 per cent decrease in the odds of death. The actual result was a 33 per cent decrease in the odds of death (Clarke and Stewart, 1995). The people involved in the trials did not know. Too often modelling involves getting a group of clinicians together in a room and asking 'what will be the outcome after 10 years, can you give us your best estimate?'

- surrogate measures are often used in modelling. Again there is considerable evidence that surrogates are very rarely validated. The CD4 count is a very poor predictor of future survival from HIV infection but it has been used in economic analyses to show that treatment is cost effective, when it is not. Cholesterol lowering is not a very good proxy for assessing the effectiveness of many drugs because of harmful side effects.

- decision analytic models are used to compare alternative strategies by comparing all possible outcomes, probabilities, costs and utilities. They can be a very positive decision aid because they are very explicit about all the probabilities and outcomes and utilities and costs. There are also now more advanced Markov models which include the transition probabilities between a finite number of different health states where there is on-going risk (Sonnenberg and Beck, 1993). The problem with decision analytic models is that there are many sources of bias and it is often difficult to know when there is bias:

  - there are many common errors in model construction, and, unless the user understands the area very well, it is difficult to find out if it is biased. Bias may be hidden in the equations;

  - there is often bias in the assumptions and poor sensitivity analysis. There may, for example, be very little information to provide accurate transition probabilities (Pettiti, 1994) and these may be assumed to be independent of previous transition probabilities. There may be huge uncertainties, and the problems of generating estimates and ranges are often ignored. The same data given to different people would generate different answers;

  - the framing of the question can sometimes lead to only one answer or to other bias;

  - filling in the data holes can introduce bias. For example, many models use the utilities ascribed to patients by clinicians, but we know from the literature that clinicians have very little idea of the experiences of patients.

  - unlike the cases of randomised controlled trials and observational studies we have not yet developed check lists to be able to critically appraise decision analytic models.

- to estimate cost-effectiveness when there are gaps in the data. The problem is that decision analytic models are very often used when there are poor data. Conversely decision analytic models are particularly good when you *do* have the input data, not when you don't. If you don't have the input data what are you modelling? It must be assumptions that come from the clinicians. However, the reason we use modelling is because we do not like accepting the implicit judgements of the clinicians. So we have models that are supposed to be scientific, but then put in all the assumptions from clinicians that we said the model was designed to replace.

Professor Sheldon went on to argue that he was not against modelling *per se* but against the way modelling is often used. He quoted Drummond (1992) as saying that economic evaluations 'with their less developed methods, may be easier to manipulate. They are frequently based on assumptions, and many evaluations, particularly those based on modelling approaches have a 'black box' feel about them...'. In addition, he quoted Udvarheyli et al (1992) who said that 'this inability to verify underlying assumptions, and the inability to assess the robustness of conclusions based on them, lead to serious questions about the reliability of study findings.' The problem is that even tiny differences in alternative strategies are enough to influence decision making.

He recognised that decisions under uncertainty need to be made, and that there is often a confusion, which Dr Luce identified, between probabilities and risk on the one hand and uncertainty on the other. Putting a probability into a model does not deal with uncertainty.

He summarised his main concerns with modelling:

- analysts have been seduced by the software, the mathematics and the ease of producing estimates. It's great fun, it's intellectually very satisfying, but that doesn't make it science;

- often the uncertainty is such that it simply becomes a mathematical codification of individuals' assumptions and biases;

> *'Analysts have been seduced by the software, the mathematics and the ease of producing estimates. It's great fun, it's intellectually very satisfying, but that doesn't make it science.'*
>
> PROFESSOR SHELDON

- there is no developed system of validation or review and thus it is very hard for the user to discriminate; it is often too complex to judge on the basis of face validity;

- few models can be validated by prediction *before use* since the reason it is being done is to *save time* or make a decision before results are available. Few of them are then tested against more reliable evidence. Events have moved on **or**

even the fact of the model is used as an excuse for not doing the further empirical work. It leads to a cycle of development of a model, prediction, implementation with little feed back.

- economists argue that modelling is used in physics and chemistry, but these scientists validate their models because they collect data and test them. We don't test these models, it's difficult to, because it will take 20 years to find the data. It is not a parallel.

> *'We have models that are explicit and scientific and then put in all these assumptions from clinicians. However, the reason we use modelling is because we don't like the implicit judgements of the clinicians.'*
>
> PROFESSOR SHELDON

- modelling is in its infancy, it may lead to biased results. It's not a criticism of modelling itself. Let's develop models, and test them. Let's try and validate them. This debate is taking place within the context of a huge battle either to sell health care products or to constrain health care costs.

THE ALTERNATIVES TO MODELLING

Professor Buxton set out the advantages of randomised controlled trials for economic evaluation:

- we can expect comparability of groups;
- double blinding reduces subjective bias;
- there is a relatively easy process for data collection. Clinical researchers may be unhappy about adding significantly to data collection forms, but the processes are in place;
- there is a familiar methodology. We know how to undertake statistical analysis. We have check lists to assure ourselves as to how well it is being done;
- it gives a common source for efficacy and economic data.

He moved on to set out the problems with randomised controlled trials for economic evaluations:

- trials often enroll a sub-set of patients;
- trials are often undertaken in atypical settings. In many health care systems those people who are busy engaging in clinical trials don't really represent typical health care providers;
- resource use may be protocol-driven in a variety of ways;
- resource use is likely to be affected by blinding;
- we have not thought enough at all about how being in a trial affects patients' health state values. We should be very much more cautious about how we can use trials to get health state valuations. Patients who are willing to enter a trial have already shown attitudes to uncertainty which many people would not share.

There are ways of improving randomised controlled trials and making them more relevant to economics. More naturalistic pragmatic trials would have:

- a normal setting;
- an appropriate choice of comparator – usually current practice;
- an extended period of observation;
- follow up for all the patients for the full period, not letting drop out from study treatment mean drop out from data collection;
- an unobtrusive trial protocol leaving more discretion for the doctor.

Dr Luce set out his assessment of the alternatives to models:

- piggy-backing health economic and other outcome research on to randomised controlled trials (RCTs). The strengths of the RCT piggy-back would be the causal inferences associated with such a design. Internally valid safety and efficacy results are produced by a well accepted scientific technique. It is standardised and a good choice for cost effectiveness under certain conditions, for instance in acute care, where the randomised controlled trial (like a sepsis trial) is mimicking what happens in the real world. The weaknesses of this piggy-back design is that it is generally very poor in terms of external validity: – there are protocol induced costs; it typically measures intermediate rather than final outcomes; the trials are expensive (although the added cost of the economics is not much); and they can be time consuming. Sometimes a randomised controlled trial is just not possible. In non-pharmaceutical applications such as surgery we do not see many randomised controlled trials. At times trials are insufficient because of, for instance, chronic conditions where we need to follow patients much longer than in the typical clinical trial.
- doing naturalistic trials, a randomised trial in the real world. Its strengths are that they embrace an experimental design, they provide

decent internal and external validity and provide more realistic outcomes. However, naturalistic trials are very expensive and very time consuming, they usually centre on intermediate rather than final outcomes, and often it is not possible to carry them out.

- using quasi-experimental observation data through chart reviews or claims analysis. Observational studies have high external validity. They use real world economic data, and are usually less expensive and less time consuming than RCTs. However they have very low internal validity. Data quality is often very poor. In particular the health outcome data is either poor or non existent.

He concluded that none of these techniques were necessarily sufficient without the use of models. Models can extend randomised controlled trials and naturalistic trials to simulate final health outcomes, community practice patterns, opportunity costs and artificial study arms. They can be used to simulate conditions other than when the actual trial took place. They are useful when randomised controlled trials are impossible or impractical; they are inexpensive, quick and flexible.

WHERE DO WE GO FROM HERE?

Professor Buxton noted that certainly in the short run there are many situations where we have got to use models. To strengthen models we should:

- be cautious about the idea that we want to make models more complex. The more complex they become the more difficulty we have in trying to establish whether we think they are plausible, where they have face validity, or indeed just to understand what is happening;

- make presentation as transparent as possible. We want to avoid black boxes. If we need complex models, we can often achieve that by incrementally adding complexity. We can start off by explaining the simple model that has intuitive appeal, and then gradually develop it in a way that shows the effect of adding complexity;

> *'We have to make models as transparent as possible... we should validate against future observation... (and) against other models.'*
>
> PROFESSOR BUXTON

- start with the 'hardest' data (based on trial use if possible). No one is suggesting that we should base a model on soft data where hard data exist, but what we need to do is use the hard data and, identify the important gaps in our knowledge;

- be conservative in making base line assumptions and be cautious in conclusions;

- validate against future observation, although the evidence that that provides comes a little bit late. However, we have not seen enough effort at validation against other models using different techniques. This is a rather useful way of assessing the plausibility of a model. We can also test a model against any available data not used to estimate its co-efficients.

> *'We can use models to carry out sensible combination of* ***reliable*** *information... and to identify better where there are gaps in our knowledge.'*
>
> PROFESSOR SHELDON

He quoted Albert Einstein's advice (Cohen and Cohen, 1980) 'that everything should be made as simple as possible but not simpler' and that seems to be the compromise that satisfies all parties. His conclusions were that:

- we need to use economic evaluation to influence the design of trials, begin to reach points where trials are more appropriate to the economic questions that we want to ask;

- an ideal economic trial would probably involve randomisation but would then be 'naturalistic';
  - random allocation to initial therapies;
  - sample size set to accommodate socio-economic variability;
  - doctor and patient not blinded;
  - all patients followed for a lengthy period.

In the meantime we have to remember that even where we base our study on an RCT we are almost inevitably going to be using some modelling.

Professor Sheldon argued for a limited role for modelling:

- to carry out sensible combination of *reliable* information on effectiveness, costs and other parameters to compare whole treatment strategies. These should be subject to standard minimum levels of sensitivity analysis;

- to identify better where there are gaps in our knowledge, to assess how important those are and to provide useful information on whether further evaluation is likely to be worthwhile.

He concluded that models:

- are particularly useful at an early stage, when we haven't got observed data. We can use the process of modelling to identify what is important to collect and what sort of sample sizes are needed;
- should be more transparent and robust;
- are not always the best way of answering the question;
- are not a substitute in the longer term for good data.

Dr Luce concluded that we need to:

- have a meeting of minds between the Bayesian decision analysist's and the classical statisticians;
- embrace modelling as a useful and essential technique to estimate the value of health interventions;
- develop methods to combine random and non-random uncertainty. We have not done enough work there and that there are very technical issues that need to be explored;
- define what is an economically meaningful difference;
- educate the FDA on the value of modelling. The FDA is far from the position that the Australian and Canadian authorities seem to be in;

> *'We need to embrace the maxim 'the perfect cannot be the enemy of the good'.'*
>
> DR LUCE

- do work on research to assess the validity of modelling, to do some a postiori analyses. There has been very little work on assessing how valid models are after the fact;
- think through what is an acceptable threshold for the acceptance of models, something analogous to the 95 per cent confidence interval used in classical statistic testing;
- embrace the maxim 'the perfect cannot be the enemy of the good' which, essentially, is what this whole thing is about.

CONCLUDING COMMENTS ON THE ISSUES RAISED

The discussion demonstrates that, whilst there is agreement about the issues, there is no consensus about the role of modelling. All participants agreed that it can be a useful tool for analysis and research but differed as to its role in assisting decision making. The views of the ISTAHC panel ranged on this issue from a wish to embrace modelling as a low cost route to decision support, through one of recognition that it is an inevitable fact of life, to a belief that it can be positively dangerous. The key points to emerge are:

- modelling has become controversial because economic evaluation is now used in decision making but decision makers differ in their willingness to accept modelling;
- models have strengths and weaknesses but so do the alternatives of addressing economic questions with randomised controlled trials, naturalistic trials, and observational studies;
- naturalistic trials could reduce the need for efficacy to effectiveness modelling, and longer trials could reduce the need for intermediate to final end-point extrapolation, but we will still not have enough 'good' trial data;
- we are therefore still likely to need models and integrative studies to support decision making;
- there is agreement that models can be used to integrate 'hard' data from different sources and to identify the key research questions;
- there is disagreement as to the extent, if any, that 'what if' models can be used to support decision making;
- transparency of assumptions in any model is crucial;
- there is no set of 'rules' for good modelling or to assess quality. Checklists for both are needed;
- a number of other research areas need to be pursued, in particular ex post verification, the potential for inter-model validation, and the possible use of 'confidence intervals' for models.

Those wishing to read more about these issues are referred to Sheldon (1996), Buxton et al (1997), Luce (1995), Rittenhouse (1996) and Mandelblatt et al (1996).

REFERENCES

Adams, M. E., McCall, N. T., Gray, D. T., Orza, M. J., Chalmers, T. C. (1992). Economic analysis in randomised controlled trials. *Medical Care*; 30: 231-243.

Buxton M. J., Drummond, M. F., van Hout, B. A., Prince, R. L., Sheldon, T. A., Szucs, T., Vray, M. (1997). Modelling in economic evaluation: an unavoidable fact of life. *Health Economics*; 6: 217-227.

Canadian Coordinating Office for Health Technology Assessment (CCOHTA), (1994). Guidelines for Economic Evaluation of Pharmaceuticals. Canada, Ottawa; CCOHTA.

Clarke, M. J. and Stewart, L. A. (1995). Obtaining data from randomised controlled trials: how much do we need for realistic and informative meta-analyses? In Chalmers and Altman (eds), Systematic Reviews. BMJ Publishing Group, London.

Cohen, J. M. and Cohen, M. J. (1980). The Penguin Dictionary of Modern Quotations (revised edition). Penguin Books.

Commonwealth of Australia, Department of Health, Housing and Community Services (1995). Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Australian Government Publishing Service.

Counsell, C. E., Clarke, M. J., Slattery, J., Sandercock, P. A. G. (1994). The miracle of DICE therapy for acute stroke: fact or fictional product of sub-group analysis? *British Medical Journal;* 309: 1677-1681.

Drummond. M. (1992). Cost-effectiveness guidelines for reimbursement of pharmaceuticals: is economic evaluation ready for its enhanced status? *Health Economics;* 1: 85-92.

Eddy, D. M., Hasselbalad, V., McGivney, W. and Hendee, W. (1988). The value of mammography screening in women under the age 50 years. *Journal of the American Medical Association;* 259: 1512-1519.

Egger, M. and Davey-Smith, G. (1995). Misleading meta-analysis. *British Medical Journal;* 310: 752-754.

Evans, R. G. (1995). Manufacturing consensus, marketing truth: guidelines for economic evaluations. *Annals of Internal Medicine* 1995; 123: 59-60.

FDA Division of Drug Marketing, Advertising and Promotion (1995). Principles for the Review of Pharmaceutical Promotion (Draft). FDA: Washington D.C.

ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group (1995). A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet;* 345: 669-685.

Kassirer, J. P. and Angell, M. (1994). The journal's policy on cost-effectiveness analyses (editorial). *New England Journal of Medicine;* 331: 669-670.

Luce, B. R. (1995). Policy Implications of Modelling the cost-effectiveness of health care technologies. *Drug Information Journal.* 29; 1469-1475.

Mandelblatt, J. S., Fryback, D. G., Weinstein, M. C., Russell, L. B., Gold, M. R., Headhorn, D. C. (1996). Assessing the effectiveness of health interventions. In Gold, M. R., Seigel, J. E., Russell, L. B., Weinstein, M. C. (ed). Cost-effectiveness in health and medicine. Oxford University Press; New York.

Montgomery, S. A., Henry, J., McDonald, G. et al (1994). Selective serotonin reuptake inhibitors: meta-analysis of discontinuation rates. *Int. Clin. Psychopharmacology;* 9: 47-53.

Peto, R. (1993). Discussion of paper by Jonathan Lomas. *Annals of New York Academy of Sciences;* 73: 237.

Peto, R., Collins, R., Gray, R. (1993). Large-scale randomised evidence: Large, simple trials and overviews of trials. *Annals of New York Academy of Sciences;* 73: 314-340.

Pettiti, D. B., (1994). Meta-analysis, decision analysis, and cost-effectiveness. Oxford University Press: Oxford.

Rittenhouse, B. (1996). Uses of models in economic evaluations of medicines and other health technologies. Office of Health Economics; London.

Schulman, K. A., Lynn, L. A., Glick, H. A. and Eisenberg, J. M. (1991). Cost effectiveness of low-dose Zidovudine therapy for asymptomatic patients with HIV infection. *Annals of Internal Medicine;* 114: 798-802.

Sheldon, T. A. (1996). Problems of using modelling in the economic evaluation of health care. *Health Economics;* 5: 1-11.

Song, F., Freemantle, N., Sheldon, T. et al. (1993). Selective serontonin reuptake inhibitors: meta-analysis of efficacy and acceptability. *British Medical Journal;* 306: 683-687.

Sonnenberg, F. A. and Beck, J. R. (1993). Markov models in medical decision making: A practical guide. *Medical Decision Making;* 13: 322-328.

Tengs, T. O. (1996). An evaluation of Oregon's Medicaid Rationing Algorithms. *Health Economics;* 5: 171-181.

Udvarheyli, I. S., Colditz, G. A., Rai, A., Epstein, A. M. (1992) Cost effectiveness and cost benefit analysis in the medical literature: are methods being used correctly? *Annals of Internal Medicine;* 16: 238-244.

This Briefing was edited by Adrian Towse and Mike Drummond.

Adrian Towse is the Director of the Office of Health Economics, London, UK.

Mike Drummond is Professor of Health Economics and Director of the Centre for Health Economics, University of York, UK.

ACKNOWLEDGEMENT OF FUNDING

The ISTAHC session was sponsored by the IFPMA Health Economics Advisory Group, with financial support from Glaxo Wellcome, Pharmacia & Upjohn, and Rhone Poulenc Rorer.