PROBABILISTIC ANALYSIS AND COMPUTATIONALLY EXPENSIVE MODELS: NECESSARY AND REQUIRED?

S Griffin, K Claxton, N Hawkins, M Sculpher, Centre for Health Economics, University of York, UK

Abstract

Decision-makers require estimates of decision uncertainty as well as expected net benefits of interventions. Recent guidance from the National Institute for Clinical Excellence (NICE) recommends probabilistic sensitivity analysis (PSA) as the appropriate way to quantify decision uncertainty. This requirement may be difficult in computationally expensive models, particularly those employing patient-level simulation (PLS). This paper discusses the importance of considering decision uncertainty, the appropriateness of PSA, and the use of PLS in NICE appraisals.

A review of published appraisals identified cases in which the model structure employed PLS. Each case was examined to see if decision uncertainty was estimated, whether PLS was necessary, and whether alternative modelling approaches were available.

One out of six case studies estimated decision uncertainty. Reasons for choosing PLS included: treatment switching; sampling from patient characteristics; dependence on elements of patient histories. Alternative approaches were demonstrated in two areas: the use of semi-Markov cohort model structure; and the use of emulators to eliminate the need for 2-level simulation.

Stochastic treatment switching and sampling baseline patient characteristics do not inform adoption decisions. Modelling patient histories is a function of the software used, and does not necessitate PLS. Where there is a non-linear relationship between patient variability and model outputs PLS is appropriate but not necessary. Where models become computationally expensive, PSA is possible using more computing power, emulators or a mathematical linear approximation. Models must provide credible estimates of net benefit and decision uncertainty. Failure to estimate decision uncertainty must be justified in terms of this dual requirement.

Introduction

Economic evaluation of health care technologies is now a necessity for informing allocation of health care resources in several countries around the world.[1] For almost every set of technologies considered, it will be necessary to combine information on costs and effects from several sources, and modelling techniques will be employed. The National Institute for Clinical Excellence (NICE) in the UK, demands decision analytic models to provide information on the expected costs and effects of alternative health care technologies. Given that there is a range of methods available for assessing cost-effectiveness, NICE has specified a Reference Case, using those methods most appropriate for informing the adoption decision, in order that all proposed technologies be assessed in a consistent manner.[2] The Reference Case now stipulates probabilistic sensitivity analysis (PSA) as the appropriate way by which the combined implications of uncertainty in all model parameters can be reflected. The guidance also indicates that when the decision model represents a (second-order) non-linear combination of parameters, probabilistic models are necessary to accurately estimate expected costs and effects, and models based on the point estimates of parameter values may provide biased results. Thus PSA should be used to provide an estimate of the decision uncertainty, i.e. the probability that the optimal treatment strategy represents the most cost-effective use of resources.

Beyond the use of probabilistic analysis, NICE does not specify the type of model to be used in the Reference Case, as this is dependent on the health care technology and the disease area under consideration. However, it does implicitly recommend those model structures that are conducive to the use of PSA. Among the technology appraisals submitted to NICE thus far, several economic evaluations have employed patient-level simulation (PLS) models, for which PSA is often computationally expensive and, as a result is not always carried out. When the new guidance comes into force, departures from the Reference Case, and therefore the use of more computationally expensive model structures that are considered to preclude PSA, will have to be justified.

This paper explores the justification for including PSA in the Reference Case. It then consider the use of computationally expensive model structures that may make the

application of PSA difficult, specifically the use of PLS. A review was performed to identify all published NICE technology appraisals that employ PLS. In each case we examine whether PSA was performed, and where this was not the case, whether alternative modelling approaches such as less computationally expensive structures or emulators were available.

Why decision uncertainty and why probabilistic analysis?

The decision whether to adopt a particular technology is based on the expected net benefit (NB) of that technology relative to appropriate comparators. Given a number of mutually exclusive treatment strategies for a particular disease area, the optimal strategy is simply the one with the highest expected NB.[3] Thus a point estimate of expected NB will suffice if the adoption decision is to be based on current evidence. However, in order to know whether the current evidence is sufficient for allocating health care resources, some measure of the decision uncertainty is required. If the decision uncertainty is large and/or the consequences of adopting a sub-optimal treatment strategy are large, the decision-maker may require further evidence on which to base the adoption decision. If it is accepted that adoption decisions should not be made without consideration of decision uncertainty, then models submitted to decision-makers must employ a method to estimate the decision uncertainty. This estimation of decision uncertainty must be done within a model that accurately estimates the net benefits of the alternative treatment options in order to adequately inform decision-makers.

One- and multi-way sensitivity analyses cannot reflect the combined uncertainty in all model parameters, and so are inappropriate for informing decision uncertainty. PSA provides a more rigorous approach by requiring that all input parameters in a model be specified as full probability distributions, rather than as point estimates, to indicate the uncertainty of the estimates.[4] Decision models comprise complex mathematical structures to synthesise evidence from multiple sources, in order to estimate the expected costs and effects of alternative treatment strategies, for patients of a particular disease area, over a specified time horizon.[5] Frequently, this will be characterised by a non-linear combination of model parameters such that the expected costs and effects cannot be calculated as a function of the expected values (point estimates) of the individual model parameters. Thus, a model that relies on the

expected value of each parameter can provide a biased estimate of the expected cost and effect of each treatment strategy. PSA can be used to accurately estimate expected costs and effects, and also to reveal the effect of the combined uncertainty in all model parameters. It also forms the first step in performing value of information (VOI) analyses: to quantify the value of acquiring more evidence, both overall and in reference to specific model parameters.[3, 6] This is important because, as well as determining the optimal treatment strategies, decision-makers such as NICE can make recommendations or provide funding for future research, and can specify a review date for when new evidence becomes available. In order to prioritise research, one must be able to identify those parameters that contribute most to decision uncertainty.

Alternatives to PSA using Monte Carlo simulation do exist in the form of analytical solutions to models and emulators. If the mathematical structure of the model is tractable, a direct solution of expected NB and possibly the associated uncertainty can be calculated. The analytical solution could take the form of a closed form solution for the posterior distribution of NB. This method may not even require computers, and would take little time. However, models structures with this property are very rare. Emulators take the form of a non-parametric statistical model of the outputs of a model, such that those outputs can be re-calculated with minimal time and computational expense when the inputs to the model are varied according to the associated second-order uncertainty.[7] The methodology of such emulators is still in development, and there are some limitations associated with their use, for example in the number of uncertain parameters that may be included.

PSA and model structure

Executing PSA within a cohort model structure is relatively simple as it entails a 1-level simulation. Within a PLS model, thousands of simulations can be required for a single run of the model using only the point estimates for the input parameters. Thus executing PSA requires a 2-level simulation where each set of probabilistic inputs is held constant while the required number of patients is simulated through the model.[7] This can make PSA an order of 1,000 or 10,000 more computationally expensive in a patient-level simulation structure as compared to a cohort structure, and it is for this reason that PSA is sometimes omitted from models employing patient-level simulation.

Review of the use of probabilistic analysis and computationally expensive model structures in submissions to NICE

From the 11th Wave of technology appraisals (starting in 2005), all submissions to NICE will be required to provide an estimate of the joint (second-order) uncertainty of all parameters.[2] The Reference Case requirement applies to the manufacturers' submissions as well as the independent assessment reports.

In order to assess the use of probabilistic analysis and computationally expensive models in submissions to NICE we reviewed the published technology appraisals to identify those where the independent academic centre utilised a model structure involving patient-level simulation for providing an estimate of cost-effectiveness. We identified common reasons for choosing a patient-level simulation rather than a cohort framework, and assess their implications for model structure. For each case study, we ascertained whether a PSA had been undertaken. In those cases where a PSA had not been undertaken, we explore the availability of alternative modelling techniques, such as less computationally expensive modelling structures, or emulators. In those cases where PSA was performed, we discuss the techniques used to overcome the computational expense of an individual patient structure.

Results

The review identified **6** assessment reports submitted to NICE where the estimates of expected costs and outcomes were based on a model structure using patient-level simulations. Details of the included appraisals are given in Table 1.

Table 1. Examples of the use of patient-level simulation in NICE assessment reports

Case study	Title	Gave justification for choice of PLS?	Estimate decision uncertainty?
1	Imatinib for the treatment of patients with unresectable and/or metastatic gastro-intestinal stromal tumours – a systematic review and economic evaluation[8]	Yes	**No**

2	The clinical effectiveness and cost effectiveness of prevention and treatment of osteoporosis[9]	Yes	**Yes**
3	The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation[10]	Yes	**No**
4	The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults[11]	Yes	**No**
5	The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy[12]	Yes	**No**
6	Coronary artery stents: rapid systematic review & economic evaluation[13]	No	**No**

PLS – patient level simulation

We identified four common reasons given for choosing a patient-level simulation structure over a cohort framework. These were treatment switching, sampling from patient characteristics, dependence on patient history (including previous events and time-in-state), and uncertainty and variability. The following review section provides a more detailed description of these reasons in the context of their use in the case studies, and assesses their implications for model structure.

Treatment switching – case study 1

Overview

This case study assessed the cost-effectiveness of imatinib for unresectable and/or metastatic gastro-intestinal stromal tumours (GIST).[8] Current guidelines at the time of the assessment recommended treatment with an initial dose of 400mg daily, with the option of proceeding to a higher dose in the event of a poor response or disease progression, and withdrawal of treatment in the absence of benefit after 8 weeks. However, due to a paucity of data, the best starting dose of imatinib and the best treatment pattern were highly uncertain. The definition of GIST was altered as recently as the year 2000, and so data on the natural history of the disease were limited. The existing economic evidence consisted of one industry submission to NICE, which was a deterministic state-transition cohort model.

The model

A simple model was constructed with only four health states: progressive disease, treatment with 400mg imatinib, treatment with 600mg imatinib and death. Patients in the imatinib treatment group began the model in with 400mg daily. Patients whose disease progressed while on treatment could move to treatment with 600mg daily, or move to the progressive disease state. Patients who failed treatment with 600mg imatinib daily moved to the progressive disease state, from which the only transitions were to remain in state or proceed to death. Patients could die at any stage in the model. Patients in the control group (i.e. no imatinib) began the model in the progressive disease state, and could remain in state or die. The cycle length was 4 weeks, and the time horizon was 10 years.

The description of the model indicates that the purpose of analysis by *first-order* Monte Carlo simulation was to allow variation in the number of patients progressing from 400mg to 600mg imatinib daily. No indication is given of time- or state-dependent transitions that might necessitate a record of patient history.

Alternative modelling methodologies

In reality, the decision to move from a dose of 400mg to 600mg would not be based on random chance. The use of first-order simulation to examine variability in prescribing practice could be replaced by examining the cost-effectiveness of alternative treatment strategies, i.e. 400mg for all patients, 600mg for all patients, or 400mg for all patients followed by 600mg for all treatment failures. The same model could then be analysed using second-order Monte Carlo simulation rather than first-order, which adds no information to the treatment decision.

Sampling patient characteristics – case study 2

Overview

This case study considers treatments for osteoporosis, specifically alendronate, etidronate, risedronate, raloxifene or teriparatide for both primary and secondary prevention of fractures.[9] The main consequence of osteoporosis is an increased incidence of fractures. An osteoporotic woman who has suffered a fracture is defined as suffering from severe osteoporosis, and is at increased risk of subsequent fractures

compared to an osteoporotic woman of the same age with no prior fracture. The existing economic evidence consisted of the 5 industry submissions to NICE. Two of these employed patient-level simulations.

The model

The model is an update of a previously constructed model by the same assessment group.[14] This in effect relaxes the time constraint normally imposed on the development of such models specifically for NICE. The structure of the model is described as being similar to a Markov model with a cycle length of 1-year, the difference being that patients are entered into the model individually, and their history is tracked. The states included in the model are osteoporotic (no fractures), hip fracture, wrist fracture, vertebral fracture, proximal humerus fracture, breast cancer, coronary heart disease and death. The model also tracks the residential status of each patient in order to assign costs associated with nursing home accommodation. The authors state the belief that reflecting the increased risk of recurrent fractures after an initial fracture, and tracking the residential status of patients in the model would be difficult in a cohort model. The choice of a patient-level simulation was also made with a view to potential future availability of data on things such as the timing of fractures.

Table 2. Relative risk of subsequent fractures for alternative initial fracture sites in case study 2[9]

	Location of Subsequent Fractures			
Prior fracture site	Hip	Vertebral	Wrist	Proximal Humerus
Hip	2.3	2.5	1.4	1.9
Vertebral	2.3	4.4	1.4	1.8
Wrist	1.9	1.7	3.3	2.4
Proximal Humerus	2.0	1.9	1.8	1.9

Alternative modelling methodologies

The secondary prevention model considers the cost-effectiveness of treatment strategies for women presenting with hip, vertebral, wrist and proximal humerus fractures. Thus the patient-level simulation is employed, in part, to track the baseline, initial fracture site for each woman as the model assumes a different baseline risk of

subsequent fractures for each different site of the presenting fracture. These baseline risks are shown in Table 2. Due to the difference in baseline risk of future events for each initial fracture site, and the fact that this characteristic is known when the treatment decision is taken, these could have been assessed as separate subgroups as it is feasible to make separate treatment recommendations for each group. This would considerably reduce the number of states required to represent that portion of the model in a cohort framework.

Dependence on patient histories – case studies 2, 3, 4 and 5

Overview

Case studies 3 and 4 examined treatments for rheumatoid arthritis.[10, 11] There is a low likelihood of long-term use for any one disease-modifying anti-rheumatic drug (DMARD), as they are not always effective, lose effectiveness over time, or cause adverse effects. Case study 5 assessed the cost-effectiveness of "newer" anti-epileptic drugs (AEDs) in children.[12] Lack of effect on seizure rate and causation of intolerable side-effects means that many patients with epilepsy are treated with a sequence of drugs.

The models

The models compared fixed treatment sequences rather than individual treatments. The discontinuation rate of each treatment was modelled as a Weibull distribution with a shape parameter not equal to one (i.e. the hazard rate was not constant). In other words, the probability of discontinuing each treatment was dependent on the time spent on that treatment. Also, the availability of future treatment options was affected by whether a patient had a toxic reaction to an earlier drug.

Alternative modelling methodologies

A separate assessment examined the use of AEDs in adults.[15] The decision problem and the events to be reflected in the model were very similar to the model of AEDs in children. As with the current case studies, time to treatment discontinuation was a function of time spent on the drug, but this was facilitated in a cohort model by employing a semi-Markov framework. This semi-Markov model was built in a statistical programming language R,[16] which can manipulate n-dimensional arrays and track the time spent in each state. This alternative model structure enabled PSA

to be undertaken to provide an estimate of the decision uncertainty, without sacrificing the time-dependent structure of the model. Representing patient histories only becomes complex in cohort models when they are built in software such as spreadsheets. In alternative software such as R or WinBUGS,[17] patient history represents an additional dimension on an array, and is easier to handle.

Uncertainty and variability

If, within a homogenous patient population, there is a non-linear relationship between a characteristic that varies between patients (such as time spent on drug) and the outcome of interest, this must be reflected in a model. The presence of such non-linearity is difficult to ascertain from the description of a model in a NICE assessment report (case studies 2 to 5 could potentially exhibit this feature), and so we present a general example instead.

Suppose the outcome of interest, cost (C), is a non-linear function of some patient characteristic (x) that varies between patients (i) according to a normal distribution with mean, μ, and variance, σ^2:

$$C_i = kx_i^2$$
$$x_i \sim N(\mu,\sigma^2)$$

We, therefore, cannot rely on the expected value of x across all patients, $E[x_i]$, and require an estimate of $E[x_i^2]$. A patient-level simulation could address this issue. However, there are two methods by which we could address this issue within a cohort framework, and thus reduce computational expense. The first of these is a 2-level simulation. This involves sampling from μ and σ^2, to predict a sample of x_i. We can then sample from the predicted x_i to provide $E[x_i^2]$ as an input to the model. The second method is to derive a linear approximation to the model. Thus we must find a mathematical function of μ and σ^2 that gives us $E[x_i^2]$. I.e. determine function G such that:

$$E[x_i^2] = G(\mu,\sigma^2)$$

A linear approximation would have the added benefit of simplifying VOI analyses, and so assist in prioritising future research. The need to account for variability in model structure under these circumstances does not negate the need to estimate decision uncertainty. Consequently, under these circumstances, both variability as well as uncertainty must be characterised in order to properly inform an adoption decision.

PSA with patient-level simulation

As can be seen in table 1, case study 2 did provide an estimate of the decision uncertainty.[9] This was made feasible through the use of an emulator by employing Gaussian processes.[7] The full model was run 80 times using different values for the inputs, in order to estimate a non-parametric relationship between the input parameters and the outputs of the model. A standard PSA would likely require 1000 runs of the model, with each run requiring 8000 patients. The non-parametric technique is called Gaussian processes, and the result is a formula for calculating the model outputs for any set of input parameters. The resulting 'model of a model' can be re-analysed instantly, allowing PSA to be undertaken with considerably less computing power and time, in this case 80,000 rather than 8,000,000 simulations.

Discussion

If it is accepted that adoption decisions should be made with consideration of the associated decision uncertainty, then we may say that models submitted to decision-makers have a dual requirement to estimate both accurate net benefits and the associated second-order uncertainty. The use of patient-level simulation is often justified with reference to the first of these. In other words the claim is made that it would not be possible to accurately estimate net benefits using a cohort framework for that particular decision problem.[18] We have identified four common reasons given for choosing a patient-level simulation, and showed that none of them preclude the use of a cohort framework.

Treatment switching can be represented easily within a cohort framework or a patient-level simulation. However, it does not directly inform the adoption decision. Where the optimal treatment sequence is not known, each potential strategy can be modelled as a relevant comparator. Sampling from baseline characteristics also does not

directly inform the adoption decision, particularly when that decision can differ between the sub-groups of patients with each characteristic. Restructuring the decision problem to separate model populations heterogeneous in terms of baseline risk into homogenous sub-groups can overcome the need to record the baseline characteristic of each patient, and hence a cohort framework will be relatively simple to implement.

This review indicates that the most common justifications for choosing a patient-level simulation are time- and state-dependent transitions. In a Markov model, these would be handled using tunnel states, and if the number of states required to represent each combination is very large, the Markov framework may become unwieldy and inefficient. In order to overcome the Markovian assumption and represent time- and state-dependent transitions within a cohort framework, it is possible to use semi-Markov processes. In order to track elements of patient history within a cohort framework, it is necessary to build the model in appropriate software, such as R. The knowledge for this may not be readily available, and so to further the use of this technique, dissemination and training are required.

In models where there is a non-linear relationship between a characteristic that varies between patients and the model output, it is necessary to account for this variability. This is distinct from the issue of baseline patient characteristics that infer a different baseline risk of subsequent events, and refers to variability within homogenous patient subgroups. For example, in case studies 3, 4 or 5, the duration of treatment, which is determined in the model, varies between patients. If there is a non-linear relationship between treatment duration and the model outputs in this instance, then the patient-level variability in treatment length must be modelled. Importantly, this does not counter the need to address second-order uncertainty. A patient-level simulation can address the issue of variability, but this can also be addressed within a cohort framework by employing a 2-level simulation. A third way to address this issue would be to find a linear approximation to the model. This simplifies the analysis greatly, and would make it easier to conduct VOI analyses.

An alternative to choosing a different model structure in order to carry out PSA is to keep the patient-level simulation and use hardware and software that reduce the run

time, or use emulators in place of a full PSA of the original model. In the updated rheumatoid arthritis case study,[11] the use of Borland Delphi[19] instead of TreeAge DATA 3.5[20] speeded up the analysis. However, in this particular case, the gain in analysis time was not enough to facilitate PSA. In the case study assessing treatment for osteoporosis,[9] Gaussian processes were employed to facilitate PSA.[7] Thus a non-parametric model of the model output, as a function of the model inputs, was estimated. This 'model of a model' could then be analysed relatively quickly to produce an estimate of the decision uncertainty. The use of such emulators is still in development, and there are currently some limitations, for example in the number of uncertain input parameters that can be included.

In the case studies presented here, the choice of a patient-level simulation model precluded the use of PSA in all but one. The reviewed models tended to rely instead on uni- and multi-variate sensitivity analyses. These will never provide a full estimate of the decision uncertainty, and are therefore less informative for decision-makers. In the epilepsy in children case study,[12] results varying according to first-order uncertainty were presented, which do not directly inform the adoption decision. If we accept the dual requirement of models to accurately estimate net benefits and the associated second-order uncertainty, then the failure to fulfil the requirement to estimate decision uncertainty must be justified in terms of more accurately estimating net benefit. In other words there may be a trade-off between the complexity of the model required to accurately estimate net benefit and the ability to perform PSA. As we have seen in the case studies reviewed here, there are often alternative modelling approaches that can be used to reduce the complexity of the model, or facilitate the conduct of PSA within a complex model structure, both of which do not reduce the ability of the analyst to correctly estimate net benefits, and as such this trade-off may not be so apparent in practice.

1. Hjelmgren, J., F. Berggren, and F. Andersson, *Health economic guidelines - similarities, differences and some implications.* Value in Health, 2001. **4**(3): p. 225-250.
2. NICE, *Guide to the Methods of Technology Appraisal*. 2004, NICE: London.
3. Claxton, K., *The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies.* Journal of Health Economics, 1999. **18**: p. 342-64.
4. Briggs, A.H., et al., *Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease.* Medical Decision Making, 2002. **22**: p. 290-308.
5. Kuntz, K. and M.C. Weinstein, *Modelling in economic evaluation*, in *Economic Evaluation in Health Care: Merging Theory with Practice*, M. Drummond and A. McGuire, Editors. 2001, Oxford University Press: Oxford.
6. Claxton, K. and J. Posnett, *An economic approach to clinical trial design and research priority-setting.* Health Economics, 1996. **5**: p. 513-524.
7. Stevenson, M., J. Oakley, and J. Chilcott, *Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis.* Medical Decision Making, 2004. **24**(1): p. 89-100.
8. Wilson, J., et al., *Imatinib for the treatment of patients with unresectable and/or metastatic gastro-intestinal stromal tumours - a systematic review and economic evaluation*. 2003. http://www.nice.org.uk.
9. Stevenson, M., et al., *The clinical effectiveness and cost effectiveness of prevention and treatment of osteoporosis*. 2004. http://www.nice.org.uk.
10. Jobanputra, P., et al., *The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation.* Health Technology Assessment, 2002. **6**(21).
11. Clark, W., et al., *The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults*. 2004. http://www.nice.org.uk.
12. Bryan, S., et al., *The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy*. 2004. http://www.nice.org.uk.
13. Hill, R., et al., *Coronary artery stents: a rapid systematic review.* Health Technology Assessment, 2004. **[in preparation]**.
14. Kanis, J., et al., *Treatment of established osteoporosis: a systematic review and cost-utility analysis.* Health Technology Assessment, 2002. **6**(29).
15. Drummond, M., et al., *A rapid and systematic review of the clinical effectiveness, tolerability and cost effectiveness of newer drugs for epilepsy in adults*. 2003. http://www.nice.org.uk.
16. *R: A language and environment for statistical computing*. 2004, R Development Core Team: Vienna.
17. Spiegelhalter, D.J., et al., *BUGS: Bayesian inference using Gibbs sampling, version 0.5, (version ii)*. 1996, MRC Biostatistics Unit: Cambridge.

18. Barton, P., S. Bryan, and S. Robinson, *Modelling in the economic evaluation of health care: selecting the appropriate approach.* Journal of Health Services Research & Policy, 2004. **9**(2): p. 110-118.
19. *Borland Delphi*. 2004, Borland Software Corporation: Scotts Valley, California.
20. *TreeAge DATA 3.5*. 2004, TreeAge Software, Inc.: Williamstown, Minnesota.