Health Technology Assessment 2004; Vol. 8: No. 27

Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis

P Tappenden, JB Chilcott, S Eggington, J Oakley and C McCabe

June 2004

Health Technology Assessment
NHS R&D HTA Programme

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis

P Tappenden,* JB Chilcott, S Eggington, J Oakley and C McCabe

School of Health and Related Research (ScHARR),
University of Sheffield, UK

* Corresponding author

Declared competing interests of authors: none

Published June 2004

This report should be referenced as follows:

Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis. *Health Technol Assess* 2004;**8**(27).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE.

NHS R&D HTA Programme

The research findings from the NHS R&D Health Technology Assessment (HTA) Programme directly influence key decision-making bodies such as the National Institute for Clinical Excellence (NICE) and the National Screening Committee (NSC) who rely on HTA outputs to help raise standards of care. HTA findings also help to improve the quality of the service in the NHS indirectly in that they form a key component of the 'National Knowledge Service' that is being developed to improve the evidence of clinical practice throughout the NHS.

The HTA Programme was set up in 1993. Its role is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The HTA programme commissions research only on topics where it has identified key gaps in the evidence needed by the NHS. Suggestions for topics are actively sought from people working in the NHS, the public, consumer groups and professional bodies such as Royal Colleges and NHS Trusts.

Research suggestions are carefully considered by panels of independent experts (including consumers) whose advice results in a ranked list of recommended research priorities. The HTA Programme then commissions the research team best suited to undertake the work, in the manner most appropriate to find the relevant answers. Some projects may take only months, others need several years to answer the research questions adequately. They may involve synthesising existing evidence or designing a trial to produce new evidence where none currently exists.

Additionally, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme is able to commission bespoke reports, principally for NICE, but also for other policy customers, such as a National Clinical Director. TARs bring together evidence on key aspects of the use of specific technologies and usually have to be completed within a limited time period.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this monograph was commissioned by the HTA Programme as project number 02/29/01. As funder, by devising a commissioning brief, the HTA Programme specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

Abstract

Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis

P Tappenden,* JB Chilcott, S Eggington, J Oakley and C McCabe

School of Health and Related Research (ScHARR), University of Sheffield, UK
* Corresponding author

Objectives: To develop methods for performing expected value of perfect information (EVPI) analysis in computationally expensive models and to report on the developments on the health economics of interferon-β and glatiramer acetate in the management of multiple sclerosis (MS) using this methodological framework.
Data sources: Electronic databases and Internet resources, reference lists of relevant articles.
Review methods: A methodological framework was developed for undertaking EVPI analysis for complex models. The framework identifies conditions whereby EVPI may be calculated numerically, where the one-level algorithm sufficiently approximates the two-level algorithm, and whereby metamodelling techniques may accurately approximate the original simulation model. Metamodelling techniques, including linear regression, neural networks and Gaussian processes (GP), were systematically reviewed and critically appraised. Linear regression metamodelling, GP metamodelling and the one-level EVPI approximation were used to estimate partial EVPIs using the ScHARR MS cost-effectiveness model.
Results: The review of metamodelling approaches suggested that in general the simpler techniques such as linear regression may be easier to implement, as they require little specialist expertise although may provide only limited predictive accuracy. More complex methods such as Gaussian process metamodelling and neural networks tend to use less-restrictive assumptions concerning the relationship between the model inputs and net benefits, and therefore may permit greater accuracy in estimating EVPIs. Assuming independent treatment efficacy, the 'per patient' EVPI for all uncertainty parameters within the ScHARR MS model is £8855. This leads to a population EVPI of £86,208,936, which represents the upper estimate for the overall EVPI over 10 years. Assuming all treatment efficacies are perfectly correlated, the overall per patient EVPI is £4271. This leads to a population EVPI of £41,581,273, which represents the lower estimate for the overall EVPI over 10 years. The partial EVPI analysis, undertaken using both the linear regression metamodel and Gaussian process metamodel clearly, suggests that further research is indicated on the long-term impact of these therapies on disease progression, the proportion of patients dropping off therapy and the relationship between the EDSS, quality of life and costs of care.
Conclusions: The applied methodology points towards using more sophisticated metamodelling approaches in order to obtain greater accuracy in EVPI estimation. Programming requirements, software availability and statistical accuracy should be considered when choosing between metamodelling techniques. Simpler, more accessible techniques are open to greater predictive error, whilst sophisticated methodologies may enhance accuracy within non-linear models, but are considerably more difficult to implement and may require specialist expertise. These techniques have been applied in only a limited number of cases hence their suitability for use in EVPI analysis has not yet been demonstrated. A number of areas requiring further research have been highlighted. Further clinical research is required concerning the relationship between the EDSS, costs of care and health outcomes, the rates at which patients drop off therapy and in particular the impact of disease-modifying therapies on the progression of MS. Further methodological research is indicated concerning the inclusion of epidemiological population parameters within the sensitivity analysis; the development of criteria for selecting a metamodelling approach; the application of metamodelling techniques within health economic models and in the specific application to EVI analyses; and the use of metamodelling for EVSI and ENBS analysis.

Contents

Glossary and list of abbreviations

Technical terms and abbreviations are used throughout this report. The meaning is usually clear from the context, but a glossary is provided for the non-specialist reader. In some cases, usage differs in the literature, but the term has a constant meaning throughout this review.

Glossary

Architecture The way in which nodes are organised within a neural network.

Basis function A function used to interpolate between data points.

Bifurcation The splitting of a group of factors into two subgroups of equal size.

Cumulative density function Denotes the probability that a random variable (e.g. *X*) has a value less than or equal to a particular value.

Design of experiments A structured, organised method for determining the relationship between factors affecting a process and the output of that process.

Elasticity A measure of the change in the value of an outcome to a change in the value of an input parameter.

Fourier frequency The oscillation frequency applied to each input parameter in a model.

Interpolation The estimation of a value between two known values.

Metamodel A statistical approximation of a simulation model.

Node A processing unit within a neural network through which data are passed.

Response surface A function of explanatory variables which represents the distribution of the expected response.

Signal-to-noise ratio A performance criterion used to determine the significance of a parameter within a model.

Spline A polynomial basis function, of which a piecewise series form a metamodel.

Training data Data used to build and calibrate a neural network metamodel.

List of abbreviations

ABN	Association of British Neurologists
AEC	actual elasticity coefficient
AROS	absolute relative overall sensitivity
CDF	cumulative density function
CI	confidence interval
CNS	central nervous system
DMT	disease-modifying therapy
DOE	design of experiments
DSS	disability status scale
EC	elasticity coefficient
EDSS	expanded disability status scale
ENB	expected net benefit
ENBS	expected net benefit of sampling
EVI	expected value of information
EVPI	expected value of perfect information
EVSI	expected value of sample information

continued

List of abbreviations *continued*

FDM	frequency domain methodology
GP	Gaussian process
GSA	generalised sensitivity analysis
HR-QoL	health-related quality of life
IFN-β	interferon-β
INB	incremental net benefit
K–S	Kolmogorov–Smirnov
LYG	life years gained
MAICER	maximum acceptable incremental cost-effectiveness ratio
MARS	multivariate adaptive regression splines
MS	multiple sclerosis
MSD	maximum separation distance
NEST	neural simulation tool
NT	net benefits
NICE	National Institute for Clinical Excellence
PCC	partial correlation coefficient
PCV	partial contribution to variance
PPMS	primary progressive multiple sclerosis
PRCC	partial rank correlation coefficient
QALY	quality-adjusted life-year
RCT	randomised controlled trial
RRMS	relapsing/remitting multiple sclerosis
RSS	risk-sharing scheme
SB	sequential bifurcation
ScHARR	School of Health and Related Research
SNR	signal-to-noise ratio
SPMS	secondary progressive multiple sclerosis
VBA	Visual Basic for Applications

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Aims and objectives

The aims of the report are: (1) to develop methods for performing expected value of perfect information (EVPI) analysis in computationally expensive models; these methodological advances will be reported and applied alongside a case study to form a clear and valuable reference source to health economists and analysts in other outcomes research organisations; (2) to report on the developments on the health economics of interferon-β (IFN-β) and glatiramer acetate in the management of multiple sclerosis (MS) using this methodological framework.

Background

The expected value of information (EVI) approach uses a decision analytic framework in order to prioritise further research through identifying those areas in which additional data collection, and hence the reduction of uncertainty, would be of most value. Value of information analysis describes the opportunity cost of uncertainty regarding a commissioning decision in terms of the probability that a suboptimal intervention is selected and the associated economic disbenefit. Further data collection is valuable if it reduces the likelihood of making the wrong decision. Step by step algorithms for performing EVPI analysis are described within the main body of the report.

Overview of case study model: the ScHARR MS model

MS is a demyelinating disease of the central nervous system. MS is the most frequent cause of neurological disability in young adults, and is typically characterised by chronic relapse and disease progression. There is no effective cure for MS; drugs known as disease-modifying therapies (the IFN-βs and glatiramer acetate) are aimed at reducing the number and severity of relapses experienced and slowing disease progression. These therapies were appraised by The National Institute for Clinical Excellence in 2001 and neither IFN-β nor glatiramer acetate was recommended for routine supply by the NHS in England and Wales. The economic analysis identified several areas of key uncertainty; however, the computational expense of the ScHARR MS model precluded the formal quantification of undertaking further research in these areas. Owing to the commercial-in-confidence evidence on the relationship between the expanded disability status scale (EDSS), costs of care and health outcomes, we have converted the original ScHARR MS model into a public domain model to facilitate estimation of the value of conducting further research on IFN-β and glatiramer acetate.

Methodological framework for performing EVPI analysis on computationally expensive models

We report a methodological framework for undertaking a comprehensive analysis of the value of perfect information for computationally expensive health economic models. This proposed framework follows a sequential logic, and identifies conditions whereby EVPI may be calculated numerically, where a one-level sampling algorithm may sufficiently approximate the more computationally expensive two-level algorithm, in addition to identifying methods for metamodelling, that is, replacing the original economic model with a statistical approximation.

This review has resolved methods for defining the number of samples required to achieve stable and unbiased EVPI estimates from the two-level EVPI algorithm and for estimating confidence intervals for EVPI estimates.

A review of the current literature identified several metamodelling approaches; the following metamodelling techniques are reviewed:

- linear regression
- neural networks
- response surface methodology (using polynomial regression)
- multivariate adaptive regression splines
- Gaussian processes/Kriging (non-linear regression).

A critique of these metamodelling methods suggests that, in general, simpler techniques such as linear regression may be easier to implement, as they require little specialist expertise although they may provide limited predictive accuracy. Conversely, more sophisticated techniques such as Kriging/Gaussian process methodology and neural networks tend to require greater specialist expertise. These more complex methods, however, tend to use less restrictive assumptions concerning the relationship between the model inputs and net benefits, and may therefore permit greater accuracy in estimating EVPI.

Applied methodology

The methodological framework was applied to the ScHARR MS model in order to estimate the value of conducting further clinical research in this area. This analysis used three separate models:

1. the original ScHARR MS model
2. a linear regression metamodel used to approximate the ScHARR model
3. a Gaussian process metamodel used to approximate the ScHARR model.

Assuming independent treatment effects, the 'per patient' EVPI for all uncertain parameters within the case study model is £8855; this represents the upper estimate for the overall EVPI. Assuming that treatment efficacies are perfectly correlated, the global per patient EVPI is £4271; this represents a lower estimate for the overall EVPI.

Due to the computation time required, it was not possible to perform two-level partial EVPI analysis for parameters using the original ScHARR model. Linear regression analysis suggested a reasonable degree of linearity between the model inputs and net benefits. A linear regression metamodel and Gaussian process metamodel were constructed in order to approximate the relationship between model inputs and net benefits. The Gaussian process model is likely to be more reliable as it is a non-linear regression technique which incorporates all possible interactions between those variables included in the simulation model.

Case study results

We estimated the relevant population for the technology over a 10-year time horizon. Assuming independent treatment effects, the global population EVPI for all uncertain parameters within the case study model is £86,208,936; this represents the upper estimate for the overall population EVPI. Assuming that treatment efficacies are perfectly correlated, the global population EVPI is £41,581,273; this represents a lower estimate for the overall EVPI. The partial EVPI analysis, calculated using both the linear regression model and Gaussian process model, clearly suggests that further research is indicated on the long-term impact of these therapies on disease progression, the proportion of patients dropping off therapy and the relationship between the EDSS, costs of care and health outcomes. Although further information on costs associated with particular EDSS states and the rates at which patients drop off therapy may be obtained through non-experimental designs such as observational studies, further useful information on the impact of disease-modifying therapies on disease progression and associated health outcomes would be most reliably obtained through a long-term randomised controlled trial which includes a direct assessment of quality of life.

Discussion

Linearity of the model

Regression analysis takes a central role in undertaking EVPI analysis via metamodelling. The main potential drawback concerns the degree of linearity between the model inputs and net benefits. If the relationship between net benefits and the parameter inputs is only weakly linear, multiple linear regression is unlikely to be useful in performing partial EVPI analysis. Conversely, if the relationship is strongly linear, it is likely that even if the expected net benefits for each treatment strategy are predicted with accuracy, the prediction error in the calculation of net benefits is likely to be magnified in the calculation of EVPI. The applied methodology clearly points towards using more sophisticated metamodelling approaches in order to obtain greater accuracy in EVPI estimation.

Where a reasonably strong linear relationship exists, the linear regression metamodel may be used in order to obtain one-level estimates of partial EVPI for all model parameters. This exercise may enable the modeller to ascertain which of the model parameters are likely to attain value and which are not, and potentially suggest an order of magnitude for this expected value. If the analyst is aware of the key parameters, it may be possible to revert back to

the original cost-effectiveness model and perform partial EVPI analysis using the correct two-level sampling algorithm for those identified parameters, and to ignore the remaining parameter set.

Although the question 'how linear is linear enough?' for use in EVPI analysis cannot be resolved using standard statistical tests, it is possible to explore the degree of approximation error resulting from a linear regression metamodel through comparing the global EVPI results calculated using the regression metamodel and the global EVPI results calculated using the original cost-effectiveness model. If the two global EVPI results are similar, this should enable the analyst to gauge the degree to which non-linearity may distort the partial EVPI estimates. If there is a considerable error between the global EVPI estimates, this should forewarn against the use of the one-level EVPI algorithm and highlight the need for non-linear methods such as Gaussian process metamodelling.

Use of metamodelling for undertaking value of information analysis

Although metamodels allow faster analysis of a problem, their use introduces an added element of uncertainty to the analysis; a metamodel can only ever approximate a system rather than fully replace it. Although many of the techniques appear similar in theory, the main difference relevant to the users of health economic models concerns the ease of use and availability of software. Many of these techniques have been applied in only a limited number of case studies, hence their suitability for use within EVPI analysis has not been demonstrated.

The suitability of these alternative metamodelling methods in performing EVPI analysis will essentially be determined by the expertise of the modeller, the time available for the project and the degree of accuracy required in the results. It is not unreasonable to postulate that when faced with a computationally expensive decision model, the general user of health economic models is primarily concerned with selecting the easiest and quickest metamodelling technique which provides reasonably accurate results. Indeed, in instances whereby the original cost-effectiveness model is approximately linear, regression metamodelling may be an adequate approach for identifying areas for investment in further research. This review has identified several classes of metamodelling technique. Although it has been possible to identify some of their more generic characteristics, these are certainly insufficient to identify one generally preferred technique or to identify a set of criteria for selecting a specific technique given specific case study characteristics.

Limitations of this study

The information currently available in the public domain on the alternative metamodelling techniques is limited. Insufficient information was available on the practical application of several of the metamodelling methodologies reviewed, hence these methods could not be confidently applied to the case study model. Further, the complexity of the ScHARR MS model means that it is infeasible to generate the partial EVPI analysis using the two-level sampling algorithm. As a result, there is no direct means of validating fully the partial EVPIs calculated using either the one-level sampling algorithm, the linear regression metamodel or the Gaussian process metamodel. Direct tests of validity have only been possible on the estimate of overall EVPI. This analysis demonstrated a high degree of linearity between sampled parameters and net benefits generated by the ScHARR MS model; this means that the exploration of the impact of non-linearities on the predictive ability of the metamodels considered and of the impact on parameter selection via importance analysis has been limited.

Further research

A number of areas requiring further research have been highlighted.

Further research indicated by the case study

The partial EVPI estimates generated using both the linear regression metamodel and the Gaussian process metamodel suggest that further research concerning the relationship between the EDSS, costs of care and health outcomes, the rates at which patients drop off therapy and in particular the impact of disease-modifying therapies on the progression of MS is required.

Inclusion of the 'relevant population' within the sensitivity analysis

Previous value of information studies have calculated the population EVPI by simply multiplying the per patient EVPI by a fixed number of patients over the lifetime of the decision. However, as the population relevant to a particular decision is itself uncertain, there remains an unresolved methodological issue concerning whether the uncertainty in the epidemiological parameters should also be accounted for within the sensitivity analysis.

Development of criteria for selecting a metamodelling approach

There exist a number of metamodelling techniques which have not been presented in this review. Methodological and case study work would be of benefit in exploring the application of the metamodelling techniques within health economic models and in the specific application to EVI analyses.

The use of metamodelling for EVSI and expected net benefit of sampling (ENBS) analysis

Due to similarities in the algorithms used, it is reasonable to suggest that metamodelling could have an instrumental role in performing EVSI and ENBS analysis for computationally expensive models.

Chapter 1
Introduction

Aims and objectives

The aims of the report are:

1. To develop methods for reducing the computation time involved in sensitivity analysis of computationally expensive models, including the application of the Gaussian process (GP) methodology. These methodological advances will be reported and applied alongside a case study to form a clear and valuable reference source to analysts in HTA academic centres, and other outcomes research organisations.
2. To report the developments on the health economics of interferon-β (IFN-β) and glatiramer acetate in the management of multiple sclerosis (MS) through the application of the methodological framework for calculating the expected value of information (EVI) for computationally expensive models.

Hence this report is intended to act as a catalyst for EVI analysis to be used more commonly within the technology assessment process in the UK, and to provide a platform for the dissemination of the two-level expected value of perfect information (EVPI) approach as an accepted technique within the economic evaluation of novel health technologies. An explicit aim of this report is to explain clearly the practical issues surrounding the implementation of the EVI methodology.

Methodological overview

The EVI approach uses a decision analytic framework in order to prioritise further research through identifying those areas in which additional data collection, and hence the reduction of current uncertainty, would be of most value. Value of information analysis describes the costs of the existing uncertainty regarding the provision of one intervention in terms of the probability that an alternative intervention should be provided and the opportunity costs resulting from the provision of the suboptimal intervention. Further data collection is thus valuable if it reduces the likelihood of making the wrong decision.

Principles of health economic decision making

Decision analysis within health economic evaluation concerns making decisions between two or more treatment options under conditions of uncertainty. A fundamental principle of decision theory is that an individual seeks to maximise his/her expected utility or payoff. Within the context of economic evaluation, policy makers acting on behalf of society are assumed to use similar criteria; health economists have put forward net benefit as the appropriate measure of this payoff. The objective function to be maximised when making healthcare commissioning decisions is thus the incremental net benefit (INB) for a given intervention, that is, the additional health benefits of an intervention after adjusting for any cost consequences.[1] In order to value the net benefit of an intervention, health outcomes, which are typically measured in terms of quality-adjusted life-years (QALYs), or in terms of 'natural' units such as life years gained (LYG), must be valued in monetary terms.

The net benefit of a treatment strategy is thus calculated as:

$$\text{net benefit } \mathrm{T}i = \lambda Q(\mathrm{T}i) - C(\mathrm{T}i)$$

where λ represents the maximum acceptable incremental cost-effectiveness ratio (MAICER) or 'willingness to pay' threshold, $Q(\mathrm{T}i)$ is the health benefit of treatment strategy $\mathrm{T}i$ and $C(\mathrm{T}i)$ is the cost of treatment strategy $\mathrm{T}i$.

Alternative methods exist for quantifying the payoff from various treatment interventions; for example, option pricing can be used to address the cost of switching between interventions. This method addresses the issue of the cost associated with the implementation of a new intervention;[2] however, these costs could be incorporated into the estimation of the net benefit, and so option pricing is not discussed further in this report.

The deterministic analysis of health economic models makes the explicit assumption that the true values of all parameters contained within the

model are known with absolute certainty. It follows, then, that the assumption of certainty within the model parameters, in turn, leads to an assumption of certainty in the model output. However, it is unlikely that all, or in fact any, of the true parameter values are known, hence health economic models are characterised by pervasive uncertainty. Probabilistic analysis of health economic models, on the other hand, assumes that the true value of each parameter is not known, but that each is described by either a parametric or empirical probability distribution. Hence, as there is uncertainty in the model parameters, so too is there uncertainty in the model output. In recent years there has been considerable emphasis on the development of appropriate methods for handling uncertainty in mathematical models, with a tendency to move away from univariate sensitivity analysis towards fuller probabilistic descriptions of uncertainty, e.g. cost-effectiveness planes, cost-effectiveness acceptability curves and distributions of incremental net benefits.[3]

Value of information analysis for prioritising further research

Value of information analysis is a natural extension of decision analysis and draws directly from existing methods for multivariate sensitivity analysis.[4] The principle aim of EVI is to quantify existing levels of uncertainty and to estimate the impact on the expected net benefit of alternative decision options through obtaining perfect information on model parameters. There has recently been a surge of interest in applying and developing value of information methods within health economic decision analysis and clinical trial design;[5–13] a detailed review of these developments was recently reported by Chilcott and colleagues.[14]

EVI permits the prioritisation of research through pursuing those research projects whereby attaining further information is expected to yield the greatest payoff in terms of expected net benefit (ENB). The ubiquitous existence of uncertainty means that there will always be a chance that we may make the wrong decision, that our adoption decision given current information may be wrong. If we had perfect information on all parameters for a given decision problem, we could be certain that we would select the optimal decision strategy. Information is valuable because it reduces the chance of making the wrong decision and therefore reduces the expected costs of uncertainty surrounding the decision.[15]

The degree to which further information attains value ultimately depends on two elements: the size of the uncertainty surrounding the decision problem and the opportunity costs associated with that uncertainty. If there is currently a very small amount of uncertainty surrounding a decision problem, then gathering further information is unlikely to revise our adoption decision, hence we are unlikely to gain any additional pay-off. However, if there is considerable uncertainty surrounding certain elements of the decision model, then acquiring exact knowledge that the parameter is at, say, the lower end of the range could lead us to choose another policy option to obtain the greatest pay-off. The value of this information is derived directly from the calculation of the opportunity cost or the benefits foregone resulting from the selection of the wrong decision alternative. The 'value' of further information is quantified in terms of the additional payoff that is obtained through switching adoption decisions. Hence information only attains value if it reduces the likelihood of making the wrong decision.

The EVPI algorithm provides an estimate of the per patient EVPI. If, for example, the per patient EVPI was large but the decision was relevant to only a small number of patients, further research may not be merited. The 'population EVPI' refers to the expected value of obtaining perfect information for the total number of patients subject to the decision, which is also subject to uncertainty, over the lifetime of the technology or impact of the decision.

Broadly, value of information analysis may be categorised into two distinct methodologies:

1. EVPI analysis
2. expected value of sample information (EVSI) and expected net benefit of sampling (ENBS).

EVPI quantifies the value of eliminating all uncertainty in the model or, in other words, of obtaining perfect knowledge of the value of an individual parameter or a group of parameters. The EVPI essentially represents a ceiling on the amount of money to invest in further research. The calculation of the EVPI for individual or groups of parameters allows the identification of specific areas whereby further information is expected to have the greatest impact on decision uncertainty. Further, through calculating the EVPI associated with individual parameters, this may also assist in determining the type and design of further research. For example, if further research is merited on, say, the costs associated with a set of

health states, this information may be adequately obtained via an observational study, thus precluding the requirement for a complex randomised controlled trial (RCT).

Alternatively, EVSI is concerned with predicting the expected reduction in uncertainty resulting from the collection of data from an additional finite sample. EVSI therefore represents the reduction in uncertainty that may be expected to result from further information from studies with predetermined sample size. ENBS incorporates the costs of this further sampling. As the sample size increases, the marginal value of additional information will diminish; thus ENBS allows the identification of the optimal sample size.

EVPI analysis may be conducted on all of the parameters within a model simultaneously, or on an individual or a subset of parameters. Calculating the EVPI for all parameters concurrently is referred to as the 'overall' or 'global EVPI'. Through calculating the overall EVPI we are essentially addressing the question 'what is the expected impact on the net benefit of a decision problem if we had perfect knowledge of the true value of all of the parameters within the model?' Alternatively, we may be interested in knowing the value of obtaining perfect information on a single model parameter (for example, the risk of an event) or a subset of parameters (for example, health outcomes for a group of health states); this is referred to as 'partial EVPI'. Although some parameters may attain little value individually, when combined they may be important; for example, the value associated with the cost of an individual health state may be negligible, whereas the value of acquiring information on the costs associated with all health states in a model may be substantial. Due to structural correlations and computational interactions between model variables, in the general case it is unlikely that a perfect direct relationship exists between the global EVPI and the sum of the EVPI for individual parameters.

Previous applications of expected value of information analysis in health economic modelling

To date, EVI analysis has been applied within a number of health technology assessment case studies, including:

- the management of symptoms presenting as possible urinary tract infections[16]
- the use of donepezil in the management of Alzheimer's disease[17]
- liquid-based cytology for cervical cancer screening[18]
- systems for the preservation of kidneys prior to transplantation[19]
- screening for inborn errors of metabolism using tandem mass spectrometry.[20]

It should be acknowledged, however, that these published analyses have adopted a variety of different mathematical algorithms to estimate the EVPI in individual model parameters. Furthermore, the majority of previous value of information analyses have been conducted on models that are less complex in structure and less expensive in terms of the time required to undertake a comprehensive EVPI analysis. Only two of the above analyses have successfully implemented the full two-level sampling algorithm.[19,20] A pilot study of the use of value of information analysis in prioritising research and development of health technologies is awaiting publication;[21] this study also uses the correct two-level sampling algorithm.

An outline of the problem: value of information analysis for computationally expensive models

Although value of information analysis has been applied with some success within health economic evaluation, methodological work has concentrated on the development of appropriate mathematical EVPI and EVSI algorithms, illustrated using relatively simple case studies. Few studies have confronted the practical problems resulting from the computational requirements of performing EVPI analysis within structurally complex models. In particular, the computational time requirements for the calculation of partial EVPI for individual model parameters may render such analysis infeasible, owing to the requirement for a two-level $n \times m$ sampling technique. However, there are a number of alternative mathematical techniques which may enable the general user of health economic models to overcome such computational problems in undertaking value of information analysis.

This report presents a case study of the cost effectiveness model developed as part of the appraisal of interferon-β and glatiramer acetate to illustrate the potential value of alternative

methods for performing value of information analysis. This case study is representative of the analytical problems arising from such computationally expensive models; a single replication of the model takes approximately 7 seconds to run using a Pentium 4 PC with 1 GB RAM, under the Windows XP operating system. The projected analysis time requirements for full value of information analysis using the School of Health and Related Research (ScHARR) MS cost-effectiveness model are shown in *Table 1*.

These calculations highlight an important issue: the number of samples required is assumed here to be 10,000 for both the inner and outer expectation of the partial EVPI algorithm (see *Box 2*). Commonly, the number of samples required to achieve stable EVPI values required is assumed to be 10,000 or 1000; this assumption is rarely justified, and indeed the methods for estimating confidence intervals from the inner level sampling have not previously been developed. This review has resolved the methodological uncertainty in this area, and an algorithm for defining the number of samples required and for determining confidence intervals for EVPI estimates is presented in the section 'Can EVPI be calculated numerically?' (p. 17).

Comprehensive partial EVPI analysis using the two-level sampling algorithm (assuming 10,000 iterations for both inner and outer level sampling) for all 128 individual parameters across all treatment strategies within the model would take approximately 2841 years. This is clearly infeasible, and value of information analysis would typically be omitted from such an assessment. Although there exist potential 'mathematical short-cuts' for performing EVPI analysis when the relationship between sampled parameter values and net benefits is linear,[1] the question remains as to how a general user of health economic models may perform EVPI analysis for computationally expensive and structurally complex models. This report therefore puts forward a comprehensive methodological framework for undertaking EVPI analysis. The merit of this framework is demonstrated through its direct application to the ScHARR cost-effectiveness model for interferon-β and glatiramer acetate in the management of multiple sclerosis.

Performing value of information analysis

This section describes the preliminary steps required for performing full value of information analysis on an existing deterministic health economic model.

Assigning probability distributions

The basic prerequisite for performing value of information analysis within any health economic model is that the model must be probabilistic. All parameters within the model should be identified and a probability distribution should be assigned to each parameter. These probability distributions should describe second rather than first-order uncertainty, that is, the current uncertainty in the mean value of each parameter as opposed to variation at the individual patient level. Although there are no prescriptive rules by which to assign distributions, and indeed such a prescription would be misguided (since the process is somewhat subjective), a recent study by Briggs and colleagues offers useful guidelines as to how distributions can be assigned to specific types of model parameter.[22]

Performing multivariate Monte Carlo sensitivity analysis

Once probability distributions have been defined for each parameter, multivariate Monte Carlo sensitivity analysis should be conducted over a large number of iterations (typically 10,000 runs) allowing all parameters to vary across their uncertain range. For each random iteration, the costs and QALYs gained for each treatment strategy should be recorded. If possible, the expected net benefit for each treatment strategy should be calculated outside this analysis so that the impact of different 'willingness to pay' thresholds may be explored.

TABLE 1 *Projected computation time for full value of information analysis using existing ScHARR cost-effectiveness model*

Type of analysis	Estimated time
Single model run (1 iteration across all seven treatment strategies)	7 seconds
Full multivariate Monte Carlo sensitivity analysis (10,000 iterations)	19.44 hours
One-level partial EVPI analysis on all groups of parameters (10,000 iterations)	8.10 days
One-level partial EVPI analysis on all individual parameters (10,000 iterations)	103.70 days
Two-level partial EVPI analysis on all groups of parameters (10,000 $\times$ 10,000 iterations)	221.97 years
Two-level partial EVPI analysis on all individual parameters (10,000 $\times$ 10,000 iterations)	2841.20 years

BOX 1 *Formulae for calculating the global EVPI*

Overall EVPI
Let
θ be the parameters for the model with defined prior probability distributions
d be the set of possible decisions or strategies
$NB(d,\theta)$ be the function of net benefit for decisions d, and parameters θ
Then:

$$\text{ENB given current information} = \max_d \{E_\theta\, NB(d,\theta)\}$$

$$\text{ENB given full information} = E_\theta\{\max_d NB(d,\theta)\}$$

$$\text{Global EVPI} = E_\theta\{\max_d NB(d,\theta)\} - \max_d\{E_\theta\, NB(d,\theta)\}$$

Algorithms for calculating the expected value of perfect information

Calculating the overall EVPI across all parameters simultaneously

The equations for calculating the global EVPI across all parameters are shown in *Box 1*.

In certain circumstances, for instance where the distribution of expected net benefit is known to be normally distributed and available, it is possible to calculate the expectations in the EVPI equation analytically. The methodological background to the analytical calculation of EVPI is well developed,[10,12] however, its use to date has been limited to numerical case studies illustrating the EVI methodology. Unfortunately, there is, as yet, no analytical solution to the calculation of partial EVPIs for individual or groups of parameters. This means that the only possible practical use of the analytical method is where the calculated overall EVPI is so low that it precludes the necessity for any further investigation. Therefore, although practically appealing, the analytical solution of EVPI methods is unlikely to be of use in addressing practical decision problems.

In lay terms, the overall EVPI is calculated numerically as follows:

- Step 1. Perform multivariate Monte Carlo sensitivity analysis for all uncertain variables within the model, and record the absolute costs and QALYs for each iteration.
- Step 2. Calculate the ENB for each iteration for each treatment strategy using the formula NB Tx = $\lambda Q(\text{Tx}) - C(\text{Tx})$
- Step 3. Calculate the average ENB for each treatment strategy over all iterations.
- Step 4. For each iteration, calculate the maximum NB across all treatment strategies.
- Step 5. Calculate the average of the maximum ENBs over all iterations.
- Step 6. Calculate the overall EVPI by taking the average of the maximum ENBs across all iterations (calculated in step 5) minus the maximum of the average ENBs across all treatment strategies (calculated in step 3).

Table 2 illustrates the calculations required for EVPI analysis for three hypothetical interventions using five random iterations.

Calculating partial EVPI for individual or groups of parameters

Although a variety of algorithms exist for the calculation of EVPI for individual or groups of parameters, a detailed critique of these approaches has already been conducted and is not necessary here.[14] Partial EVPI analysis for individual or subsets of parameters requires a two-level algorithm, which uses two nested levels of Monte Carlo sampling over the plausible ranges for both the parameters of interest, and the remaining uncertain parameters. *Box 2* presents the equations for calculating the partial EVPI for individual or groups of parameters in a model.

In lay terms, the steps required to calculate the partial EVPI for parameters using the two-level sampling algorithm are as follows:

- Step 1. Sample once from the parameter of interest (θ_i) and hold that parameter constant at its sampled value. If the analysis is for a group of parameters, all parameters of interest (θ_i)

BOX 2 *Equations for calculating partial EVPI for parameters using the two-level algorithm*

Partial EVPI for subsets of parameters
Let
θ_i be the parameters interest for partial EVPI
θ_{-i} be the other parameters
Then:

$$\text{ENB given perfect information on } \theta_i = E_{\theta_i}\{\max_d(E_\theta\, NB(d,\theta)\,|\,\theta_i)\}$$

$$\text{Partial EVPI obtaining data on } \theta_i \text{ only} = E_{\theta_i}\{\max_d(E_\theta\, NB(d,\theta)\,|\,\theta_i)\} - \max_d\{E_\theta\, NB(d,\theta)\}$$

TABLE 2 *Illustrative example of EVPI calculation for three hypothetical interventions*

	Monte Carlo results						EVPI calculation			
Iteration no.	**Cost T0 (£)**	**Cost T1 (£)**	**Cost T2 (£)**	**QALYs T0**	**QALYs T1**	**QALYs T2**	**ENB T0 (£)**	**ENB T1 (£)**	**ENB T2 (£)**	**Maximum ENB (£)**
1	170,608.51	161,571.61	169,534.60	14.28	14.26	14.45	257,791	266,228	263,965	266,228
2	83,818.37	77,907.92	79,914.31	13.17	13.38	13.64	311,282	323,492	329,286	329,286
3	96,834.06	89,933.29	97,557.92	8.85	8.46	9.3	168,666	163,867	181,442	181,442
4	104,826.72	94,382.59	105,785.36	11.62	11.37	11.3	243,773	246,717	233,215	246,717
5	113,070.80	96,094.69	109,543.76	10.35	10.41	10.55	197,429	216,205	206,956	216,205
Mean	–	–	–	–	–	–	235,788	243,302	242,973	247,976
EVPI										4,674

MAICER = £30,000.
Note: in this instance, T1 is the optimal choice.

should be sampled once and held constant at their sampled value.
- Step 2. Let all other model parameters **not of interest** (θ_{-i}) vary according to their prior uncertainty.
- Step 3. Run the model over n iterations (say n = 10,000) and record the costs and QALYs accrued for each treatment strategy.
- Step 4. Calculate and record the EVPI as shown in *Box 2*.
- Step 5. Repeat steps 1–4 for all individual or groups of parameters of interest.
- Step 6. The partial EVPI for the parameters of interest is simply the average of the EVPIs recorded over m iterations.

Under certain conditions [see the section 'Is the model linear?' (p. 18)], a one-level algorithm may be used in order to reduce computation time (*Box 3*).

In lay terms, one-level partial EVPI for parameters may be calculated using the following algorithm:

- Step 1. Let all parameters of interest (θ_i) vary according to their prior uncertainty.
- Step 2. Hold all parameters not of interest (θ_{-i}) at their mean value.
- Step 3. Run the model over n iterations (say n = 10,000 iterations) and record the costs and QALYs accrued for each treatment strategy.
- Step 4. Calculate the NB for each random iteration for all individual treatment strategies using the equation NB Tx = λQ(Tx) – C(Tx).
- Step 5. Calculate the average ENB for each treatment strategy.
- Step 6. For each iteration, record the maximum NB (i.e. the greatest net benefit across all treatment strategies for each iteration).
- Step 7. Calculate the average maximum ENB of all treatment strategies.
- Step 8. Calculate the overall EVPI by taking the average of the maximum ENBs across all iterations minus the maximum of the average ENBs across all treatment strategies.

It is noteworthy that the main differences between the global EVPI and one-level partial EVPI algorithms concern whether model parameters are allowed to vary or whether the parameters are held constant at their mean value. Therefore, it is possible to construct a generic model structure by which to record the NBs of each treatment strategy for each Monte Carlo iteration, such as the basic structure shown in *Table 2*. This also suggests that a common Visual Basic for Applications (VBA) subroutine may be used alongside EXCEL to perform the Monte Carlo sampling.

Identifying the relevant population for EVPI analysis

The algorithms provided above estimate the EVPI for an individual patient. The 'per patient' EVPI does not, however, provide information concerning the total expected value of obtaining 'perfect' information, i.e. the expected value of obtaining perfect information for the number of patients subject to the decision over the lifetime of the technology or impact of the decision. Consider two hypothetical decision problems, denoted 'decision a' and 'decision b', with associated per patient EVPIs of £5000 and £2000 respectively. On the basis of the 'per patient EVPIs', 'decision problem a' would be prioritised over 'decision problem b', that is, money would be invested in 'decision problem a.' However, if 'decision problem a' is associated with a relevant population of 100 patients and 'decision problem b' is associated with a relevant population of 1000 patients, and the effective lifetime of the decision is 5 years for each decision problem (discounting at 3% per annum), the population EVPIs would be

'Decision problem a' = £2,358,549
'Decision problem b' = £9,434,197

Hence there is greater value in obtaining perfect information on 'decision problem b' than 'decision problem a'. Similar considerations would also hold

BOX 3 *Formulae for calculating partial EVPI for parameters using the one-level algorithm*

Partial EVPI for subsets of parameters
Let
θ_i be the parameters of interest for partial EVPI
θ_{-i} be the other parameters
Then:
ENB given perfect information on $\theta_i = E_{\theta_i}\{\max_d(E_\theta NB(d,\theta)\,|\,\theta_i)\}$

Partial EVPI obtaining data on $\theta_i = E_{\theta_i}\{\max_d NB(d,\theta_i\,|\,\theta_{-i} = \overline{\theta_{-i}})\} - \max_d\{E_\theta NB(d,\theta)\}$

in comparing interventions where the lifetime of a decision was likely to vary between technologies, for instance in considering fast-changing technologies or competing products coming to market. It is therefore necessary to estimate the likely population relevant to a decision problem over the lifetime of the decision. This calculation essentially requires an estimate of the number of patients eligible for treatment with the intervention, including estimates of incidence and prevalence together with an estimate of the likely lifetime of the decision.

Structure of the report

The structure of the report is as follows:

Chapter 2 Development of the ScHARR multiple sclerosis cost-effectiveness model

This chapter provides an overview of the economics of IFN-β and glatiramer acetate and outlines the development of the ScHARR cost-effectiveness model required to undertake value of information analysis. The chapter describes the steps taken to convert the ScHARR cost-effectiveness model developed as part of the National Institute for Clinical Excellence (NICE) appraisal process, which was held as commercial-in-confidence, to a 'public domain' model.

Chapter 3 Methodological framework for undertaking EVPI analysis

Chapter 3 presents a methodological framework which addresses those issues which may be pertinent when attempting to perform value of information analysis for computationally expensive models. This chapter reports on the issues surrounding the immediate feasibility of performing EVPI analysis for complex models and considers alternative approaches and the circumstances under which alternative methods may be considered appropriate and robust. We present a review of metamodelling techniques which may be used to replace an existing health economic model with a statistical approximation. The role and value of each of these methodologies are reported together with salient issues that should be considered during the modelling process. As some of these approaches are restricted in terms of the number of parameters that can be included in the metamodel, we also present a systematic review of alternative approaches to ranking model parameters according to their relative importance.

Chapter 4 Applied methodology: EVI analysis for computationally expensive health economic models

This chapter reports on the sequential application of the methodological framework outlined in Chapter 3 to the case study ScHARR MS cost-effectiveness model. The applied methods are described and conclusions are drawn as to their validity, accuracy and robustness for this particular case study. The potential reduction in computation time, ease of application, existence of parametric restrictions and requirements for specialist expertise are also presented for each method.

Chapter 5 The expected value of perfect information for interferon-β and glatiramer acetate in the management of multiple sclerosis

This chapter summarises the results and conclusions of the EVPI analysis on IFN-β and glatiramer acetate in the management of MS. The overall EVPI results are presented alongside the partial EVPI results for individual parameters within the model.

Chapter 6 Discussion and conclusions

This chapter summarises the results and conclusions drawn from the methodological aspects of this study. It presents a discussion of key issues in the application of the methodological framework for undertaking EVI analysis and identifies a number of areas requiring further research.

Chapter 2

Development of the ScHARR multiple sclerosis cost-effectiveness model

Introduction

This chapter provides a brief description of the case study problem and details the history of economic evaluations of IFN-β and glatiramer acetate for MS. It also describes the necessary development of the ScHARR cost-effectiveness model into a 'public' model. In order to allow transparency and clarity of the value of information methodology, the expanded disability status scale (EDSS) cost and utility model data held as commercial-in-confidence within the original ScHARR model have been replaced with public domain estimates. Inevitably, there is a small difference between the base-case results for the public domain model and those originally reported by Chilcott and colleagues.[23]

Background to assessments of interferon-β and glatiramer acetate in the management of multiple sclerosis

Clinical background to MS

MS is a demyelinating disease of the central nervous system (CNS). MS is the most frequent cause of neurological disability in young adults and is typically characterised by chronic relapse and disease progression.[24] Evidence suggests that MS results from an autoimmune response, resulting in inflammation, demyelination and axonal loss.[24] Three commonly used categories of MS have been defined: relapsing/remitting MS (RRMS); secondary progressive MS (SPMS); and primary progressive MS (PPMS).[24]

MS is approximately twice as common in women than men. The prevalence of MS in England and Wales is conservatively estimated to be between 58,000 and 63,000 people.[25] The annual incidence of MS in England and Wales is estimated to be around 3.8 per 100,000 people. It is estimated that around 30% of these individuals may be eligible for treatment with the IFN-βs and glatiramer acetate.[25]

Disease progression is typically measured in terms of impairment and disability using Kurtzke's EDSS,[26] an ordinal scale ranging from EDSS 0 (normal neurologic examination) to EDSS 10 (death due to MS). The EDSS is presented in Appendix 1. Up to EDSS 3.5, the scale measures neurological impairments that are likely to have limited if any impact upon the activities of daily living. EDSS scores between 4.0 and 5.5 reflect ambulatory limitations for distances up to 500 m and the use of mobility aids. For scores over EDSS 6.0, patients will require a wheelchair. The progression to SPMS normally takes place over the EDSS range 2.5–4.5. Disability progression is associated with permanent reductions in quality of life (QoL) and increases in the cost of medical management.[27]

Conventional management of MS typically consists of drug therapy, physiotherapy, psychiatric and social support and disability aids; these interventions may provide symptomatic relief but have no impact upon the underlying nature of the disease. There is no cure for MS; however, IFN-β and glatiramer acetate may alter the clinical course of the disease through slowing disease progression and reducing the number and severity of relapses experienced. The use of these disease-modifying therapies (DMTs) is subject to eligibility criteria as defined by the Association of British Neurologists (ABN). These criteria are shown in *Box 4*.

Economic evaluations of disease modifying treatments for MS

The cost-effectiveness of DMTs in the management of MS has been the focus of significant attention for much of the last 10 years.

BOX 4 *ABN guidelines for the use of IFN-β and glatiramer acetate in multiple sclerosis*[28]

1. Able to walk independently
2. At least two significant relapses in the last 2 years
3. Adult age group (18 years or older)
4. There are no contraindications

Adapted from: http://www.theabn.org.

An Executive Letter in 1994, from the NHS Executive, informed all health authorities that they must manage the introduction of these new drugs with respect to local resource implications.[29] The annual cost of these therapies, which is in excess of £5000 per patient per year, combined with the length of time for which patients might continue to take these drugs, meant that funding these therapies would entail extremely large increases in expenditure. The scale of the increase was generally considered to be beyond the capacity of existing health authority budgets. If the scale of the benefits could be demonstrated to be efficient in terms of the cost per LYG or QALY gained, then increases in total funding or the sacrifice of other healthcare activities might be justified. Cost-effectiveness analysis was hence required to establish whether the funding of these new therapies represented an efficient use of resources.

To date, there have been numerous attempts to estimate the cost-effectiveness of DMTs for MS in the form of both independent and company-sponsored evaluations. Existing analyses have produced a range of cost-effectiveness estimates from in excess of £1 million per QALY gained to cost saving.[30–32] Significant flaws in the modelling of natural history, efficacy, discontinuation of therapy, mortality and the treatment of uncertainty mean that none of these estimates can be considered robust.[33] In general, those models which produced very high cost-effectiveness estimates tended to have shorter time horizons (less than 10 years) or assumed that all benefit ceased when the patient stopped therapy. Models that assumed long time horizons and sustained benefit after the cessation of therapy produced economically attractive cost-effectiveness estimates.

Having reviewed the existing evidence, NICE commissioned a new model from a consortium of universities to address explicitly the limitations identified in the review of existing models. It is this model, hereafter referred to as the ScHARR model, which is used for the research presented in the remainder of this report.

Current guidance on the use of interferon-β and glatiramer acetate for MS in England and Wales

As a result of the scientific and non-scientific evidence made available to the Appraisal Committee at NICE, on the basis of their clinical and cost-effectiveness neither IFN-β nor glatiramer acetate was recommended for the treatment of MS in the NHS in England and Wales.[34]

Following the dissemination of the NICE guidance, the Department of Health entered into price negotiations with Serono, Schering, Biogen and TEVA/Aventis, the manufacturers of IFN-βs and glatiramer acetate. The result of these negotiations was the development of a risk-sharing scheme (RSS), designed to monitor the cost-effectiveness of the four DMTs in the management of MS. The scheme involves the detailed monitoring of a cohort of patients to collect further data on the impact of DMTs on disease progression and the severity and frequency of relapses experienced. The interventions are hence available to all patients with RRMS and those with SPMS in which relapses are the dominant clinical feature, given their eligibility according to the ABN guidelines.[28] The monitoring process and associated price adjustments are expected to continue for 10 years.

Synopsis of the existing ScHARR MS cost-effectiveness model

A brief summary of the ScHARR cost-effectiveness model is presented below. A more detailed description of the model was reported by Chilcott and colleagues[23] and by Tappenden and colleagues.[35]

Model structure

The ScHARR model uses the state transition methodology to simulate the natural history of MS over the EDSS[26] across RRMS and SPMS. The model estimates the cost-effectiveness of four products licensed for RRMS. These are (in units per week):

- 6 MIU IFN-β-1a (Avonex®)
- 8 MIU IFN-β-1b (Betaferon®)
- 20 mg glatiramer acetate (Copaxone®)
- 22 μg IFN-β-1a (Rebif®)
- 44 μg IFN-β-1a (Rebif®).

The model also estimates the cost-effectiveness of 8 MIU IFN-β-1b (Betaferon®) for the treatment of SPMS. The outcome measure used to assess cost-effectiveness is the cost per QALY gained. Patients progress through the model according to instantaneous hazard rates derived from a 25-year study undertaken in London, Ontario, Canada.[36] Costs and utilities are applied directly to the state populations within each of the health states over each model cycle. During any particular model cycle, patients may also experience relapse, whereby a disutility is applied. For those patients receiving DMT therapy, a disutility is also applied

to account for the experience of treatment-related side effects. The model uses an annual cycle length over a 20-year time horizon. The transitions possible during any model cycle are shown in *Figure 1*. Patients who drop off therapy remain in the same EDSS state but transit to the conventional management quadrants of the matrix.

Model assumptions

The assumptions made in constructing the model favour the novel therapies within the analysis.

- Transitions within the model are assumed to be progressive only. For example, a patient in EDSS 4.5 in the current model cycle could not regress back to EDSS 4.0 during a subsequent model cycle.
- A sustained effect of treatment on both progression and relapse beyond the trial duration was modelled. Any patient who discontinues therapy subsequently progresses according to natural history rates but retains any benefits received at no additional cost of therapy. Thus, on the EDSS, these patients never 'catch up' with those patients who only receive conventional management.
- Due to the paucity of evidence concerning the long-term efficacy of any of these therapies, the effects of treatment are assumed to be fixed and did not deteriorate or increase over time.
- The annual relative risk of 'all-cause' mortality for the MS cohort is assumed to be the same as a normal healthy population, minus the MS death observed in the natural history cohort.
- Patients started treatment according to ABN guidelines and are treated until they reach EDSS 7.0 or drop off therapy.

Within the base-case scenario, patients start treatment according to the ABN guidelines[28] and are treated until EDSS 7.0. Patients enter the

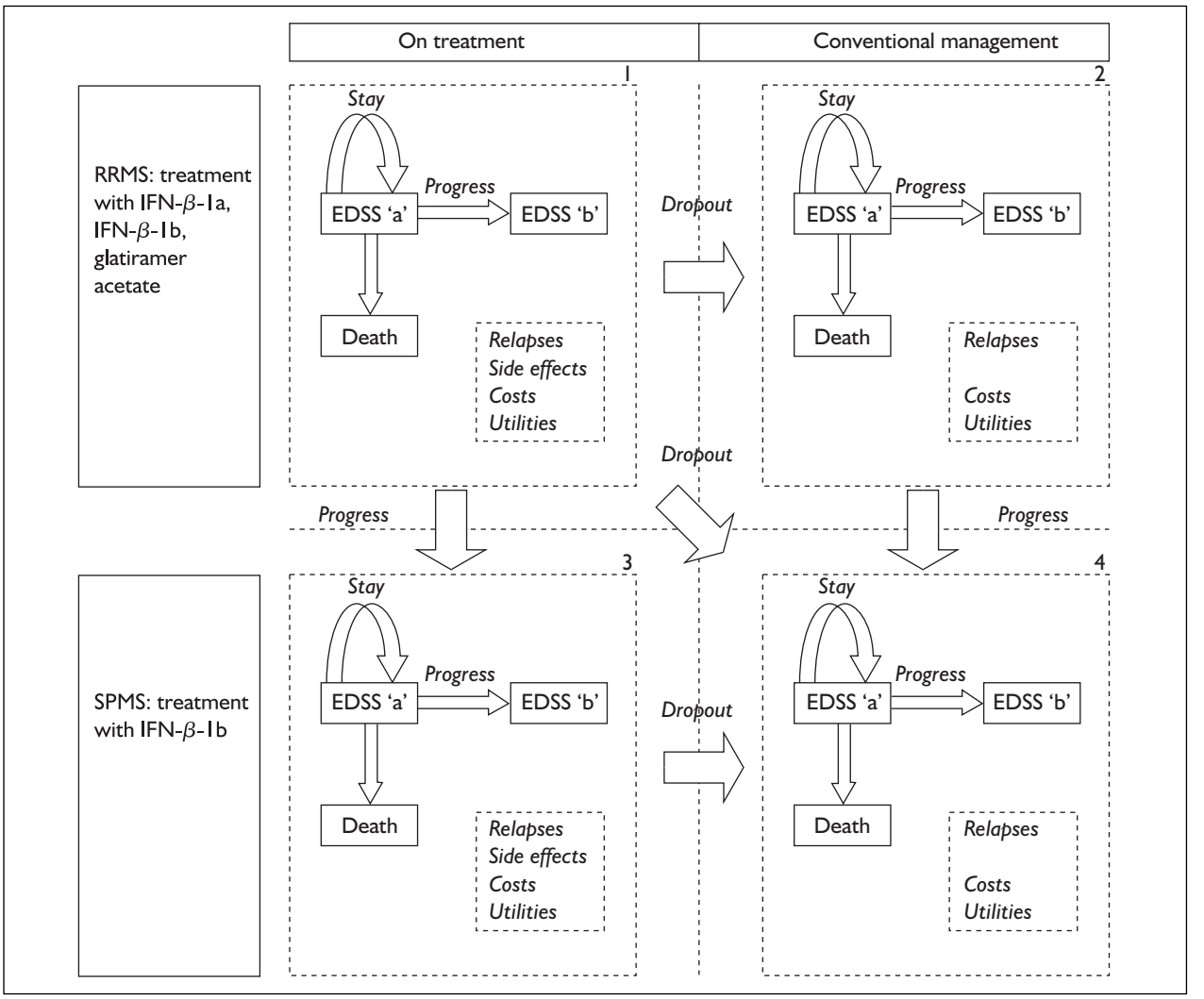

FIGURE 1 *Progression diagram for the ScHARR model*[23]

TABLE 3 *Base-case results for existing ScHARR model*[23]

Treatment strategy	Per patient results		Marginal results		Cost per QALY (£)
	Costs (£)	QALYs	Costs (£)	QALYs	
T1 IFN-β-1-a (Avonex, Biogen)	111,954	10.20	43,500	1.03	42,041
T2 IFN-β-1-a 22 μg (Rebif, Serono)	112,982	9.89	44,529	0.73	60,963
T3 IFN-β-1-a 44 μg (Rebif, Serono)	130,949	10.03	62,496	0.87	71,732
T4 IFN-β-1-b 8 MIU: treating RR (Schering)	101,726	9.83	33,272	0.67	49,664
T5 Glatiramer acetate (Copaxone, TEVA)	101,273	9.50	32,820	0.34	97,636
T6 IFN-β-1-b 8 MIU: treating RRMS and SPMS (Schering)	107,022	10.03	38,569	0.87	44,390
T0 Conventional management	68,453	9.16			

model aged 30 years. Costs and health benefits are discounted at 6 and 1.5%, respectively. Given the base-case assumptions, the cost per QALY gained ranges between £42,000 and £98,000, as shown in *Table 3*.

Conversion to a public domain cost-effectiveness model

Relationship between level of MS disability, cost and QoL

The original modelling work revealed numerous areas of considerable uncertainty concerning the natural history of MS and, hence, the impact of DMTs on the long-term clinical course of the disease. Whilst the assessment was based upon the highest quality evidence available at the time, there was a noticeable paucity in evidence, particularly concerning the impact of these therapies beyond the duration of the existing clinical trials. Although an individual with MS may live for up to 40 years,[24] the existing clinical trials of disease-modifying therapies in MS were of between 2 and 3 years in duration.[37–41] Although clinical evidence suggested that treatment with IFN-β and glatiramer acetate delays disability progression, the long-term effects of treatment on disability following cessation of therapy cannot be reliably predicted on the basis of the short-term evidence from the clinical trials.

In Chapter 4, the EVI methodology is applied to the ScHARR model in order to identify whether further clinical research is merited and, if so, for which parameters within the model data collection would be of most value.

In order to ensure that the methods described here remain transparent, it was necessary to convert the existing commercial-in-confidence model used within the NICE appraisal[23,35] to a public domain model which may be reported freely. As a result, the cost and utility data used within the original MS model have been replaced with estimates derived from published studies.

There is increasing evidence of robust relationships between an individual's EDSS and both the costs of managing their condition and the QoL associated with that health state.[27,42] In the original version of the ScHARR model, a single independent QoL weight (utility) and a cost was specified for each EDSS state in the model. However, the model did not recognise the correlation between EDSS states for either utilities or costs (i.e. as costs increase, utilities decrease in a systematic pattern reflecting the change in the underlying clinical condition). We replaced these values with functions for costs and utilities to ensure that these correlations were reflected in the model and, potentially, improved the efficiency of the model estimation processes.

The specification of the functions drew from data in the literature,[43–45] our own experience of analysing cost and QoL data in MS and our knowledge of methodological issues around cost and QoL assessment in chronic disabling conditions. On the last point, we paid particular attention to issues around data collection from severely disabled individuals, that is, patients in EDSS states 8.0 and above.

QoL assessment

It is highly unlikely that severely disabled individuals would be able to complete a QoL assessment such as the EQ-5D, hence such data are unlikely to be available. It is, however, possible to say with confidence that individuals at EDSS 9.5 would fulfil the criteria for the worst health state in the EQ-5D classification. Hence we can fix the QoL for this state (one end of the EDSS utility function) at –0.594. Further, our experience of analysing this type of data established that the shape of the function was consistently of the form illustrated in *Figure 2*.

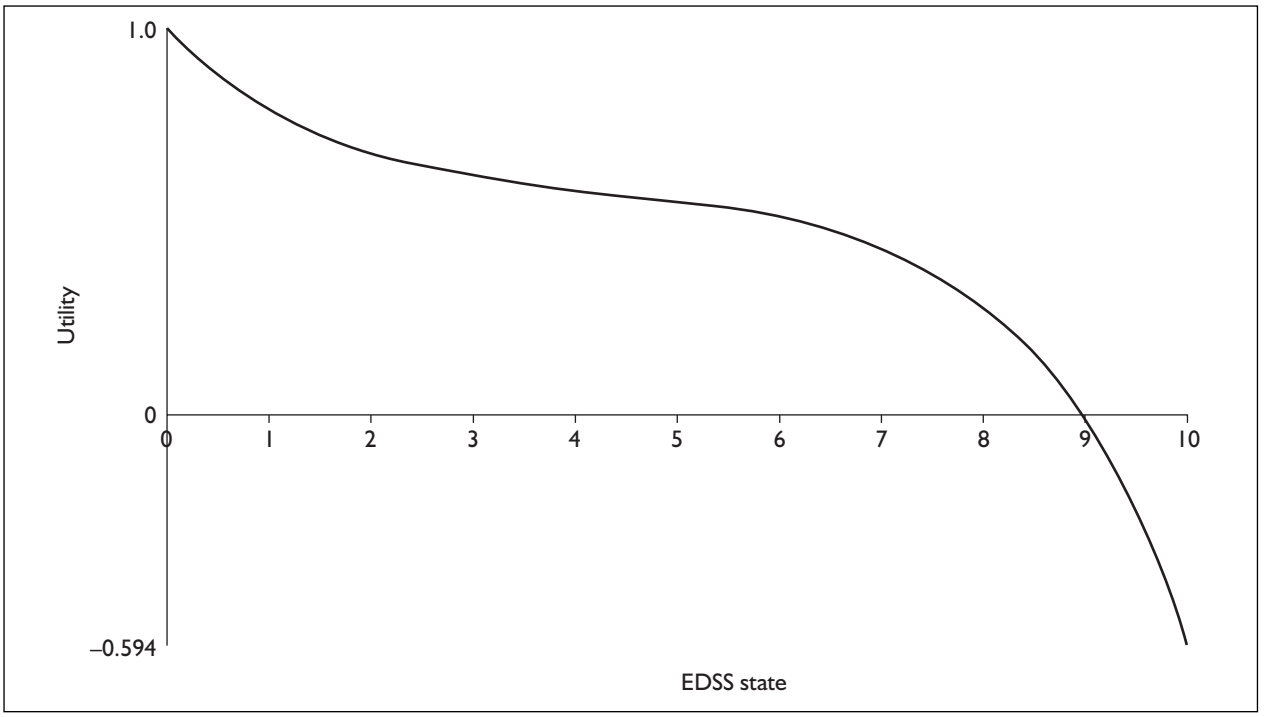

FIGURE 2 *Illustration of QoL–EDSS function*

Although there is some variation in the published QoL estimates, it is not large. In order to specify the function, it was necessary to identify the QoL value at the point for EDSS 6.0, where the function commences its sharper downward trajectory. Parkin and colleagues[44] reported a value of 0.49 for this state. As stated above, we were concerned that as the disability increases, empirical estimates will be affected by the ability or willingness of disabled individuals to complete QoL questionnaires. We therefore believe that mean values are likely to be overestimates of the true mean value for the health state. We therefore adjusted Parkin and colleagues'[44] estimate slightly downwards to reflect this belief. We set the QoL for EDSS 6.0 at a value of 0.47.

Cost of care

As with QoL, there is a clear relationship between an individual's EDSS and the cost of managing the disease. A recent review of the published literature by Patwardhan and colleagues[42] found the relationship between each EDSS state and its associated cost relative to EDSS 1.0 was of the form shown in *Figure 3*. This relationship was similar between all studies identified within the review. As with QoL, there are few data for patients in states beyond EDSS 7.5. This reflects the problems of collecting data on these patients.

As EDSS 9.5 is equivalent to complete dependency, we set the annual cost for this state equal to the cost of annual hospitalisation for a medical ward, using the 2002 NHS reference cost value.[46] The cost of management in the initial EDSS state was then fixed as the cost of an annual outpatient neurology appointment; this sets the cost of confirmation of diagnosis outside the model. Again, this cost was taken from the NHS reference cost database.[46]

As with the cost function, it was then necessary to specify a value for EDSS 6.0 in order to allow the function to be parameterised. Having chosen an annual cost that was consistent with Patwardhan and colleagues' review of published estimates (£1679),[42] we then fitted an exponential function between the fixed points of EDSS 0 and 9.5.

Correlation between treatment efficacies

In supporting commissioning decision-making, the original MS health economic model focused primarily on generating central estimates of cost-effectiveness or net benefits with only a partial analysis of uncertainty. In considering the implications for further research requirements, a fuller handling of uncertainty is required. One specific area of uncertainty is the possible existence of correlation between the efficacy of the different treatment options. Although differences exist being the products being considered, there are also marked similarities, and in these circumstances some level of correlation between treatment efficacies must be expected.

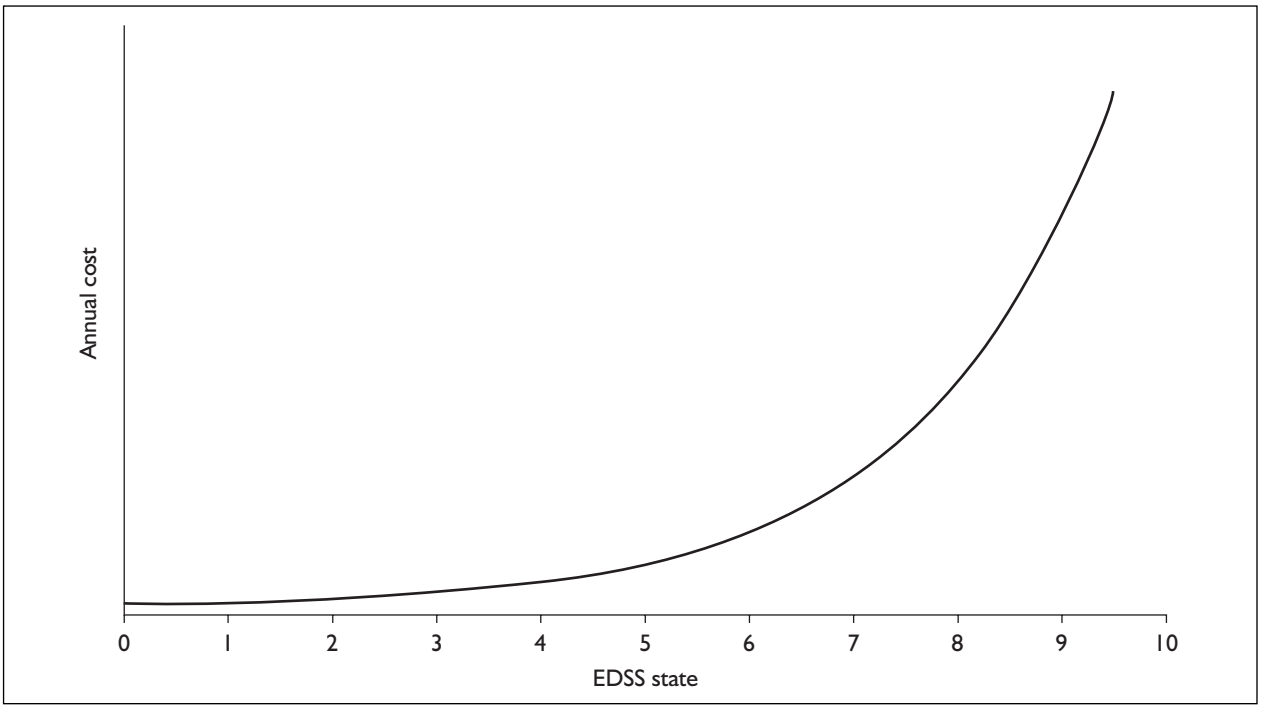

FIGURE 3 *Illustration of the shape of the EDSS–cost function*

TABLE 4 *Base-case results using revised EDSS cost and utility estimates alongside RSS drug price*

Treatment strategy	Per patient results		Marginal results		Cost per QALY (£)
	Costs (£)	QALYs	Costs (£)	QALYs	
T1 IFN-β-1-a (Avonex, Biogen)	128,997	10.68	39,874	1.02	39,277
T2 IFN-β-1-a 22 μg (Rebif, Serono)	121,966	10.36	32,843	0.69	47,318
T3 IFN-β-1-a 44 μg (Rebif, Serono)	129,363	10.51	40,240	0.84	47,828
T4 IFN-β-1-b 8 MIU: treating RR (Schering)	120,788	10.30	31,665	0.63	49,973
T5 Glatiramer acetate (Copaxone, TEVA)	114,956	9.94	25,833	0.28	92,279
T6 IFN-β-1-b 8 MIU: treating RRSM and SPSM (Schering)	126,148	10.47	37,026	0.80	46,097
T0 Conventional management	89,123	9.66			

In order to include efficacy correlation within the model, it would be necessary to handle the set of six treatment efficacies as a multivariate normal distribution and to incorporate an uncertain covariance matrix into the model. Sampling of this multivariate distribution would then be facilitated by sequentially sampling a series of standardised normal distributions and linearly transforming these samples using the Cholesky square root of the covariance matrix. This situation is further complicated by the necessity also to sample the covariance matrix in order to capture the uncertainty in the correlations between treatments.

In a practical context, it should also be noted that there is a complete absence of quantitative information on the correlations between all treatments. The only option therefore would be to use subjective judgement in defining distributions for the correlation terms.

Given these practical difficulties, it was decided to take two approaches to the analysis of EVPI for the model:

1. Include all treatment options but assume independence in treatment efficacy. This will give an upper estimate to the overall EVPI.
2. Consider a single drug treatment option, that is, that with the highest net benefit, compared with conventional management. The results provided by this analysis will be equivalent to assuming a perfect correlation between treatment efficacies since the rank ordering of ENBs will be maintained. This analysis will therefore provide a lower estimate for the overall EVPI.

Undertaking EVPI analysis within the ScHARR cost-effectiveness model

The inclusion of revised cost and utility estimates for each EDSS health state together with the current adjusted prices for the Department of Health RSS gives the results for the base case scenario shown in *Table 4*.

The application of EVPI analysis is, in general, relatively straightforward, even within structurally complex models such as the ScHARR cost-effectiveness model. Model parameters were identified and probability distributions were assigned to each parameter as part of the original modelling work conducted as part of the NICE assessment.[14,35] A description of the model parameters contained within each group is given in Appendix 2. Following the incorporation of public cost and utility estimates, revised distributions were assigned to these parameters using beta, log-normal and uniform distributional forms. This model was used as the basis for undertaking the value of information analyses using the methods detailed in Chapter 3.

Chapter 3

Methodological framework for undertaking EVPI analysis

Analytical overview

Figure 4 presents an outline flowchart for deciding on a feasible approach to calculating the EVPI in analysing a decision analytic health economic model. The methodological discussion in this report is structured around this framework. Thus specific methodological issues underlying each node in the framework are identified and addressed within each respective section. This framework is presented as a sequential approach to methods available for undertaking EVPI analysis; however, many unresolved methodological issues exist, hence it should not be considered as a prescriptive algorithm.

Can EVPI be calculated numerically?

When using the two-level Monte Carlo algorithm to estimate partial EVPIs, it is necessary to determine how many samples are needed for the inner and outer expectations. For a fixed inner sample size, it is relatively straightforward to determine the outer sample size; once the algorithm has been run for a moderate sample size (say 50), confidence intervals can be constructed for estimates using any outer sample size. The difficulty is in determining a suitable inner sample size without having to run the model a very large number of times.

An inner sample size that is too small will produce a bias in the partial EVPI estimate. The following algorithm is used to obtain an approximate estimate of the size of the bias for any inner sample size. We give an outline of the algorithm in *Box 5* and full technical details are given in Appendix 3. This method still requires a fairly large number of model runs (~600). However, these model runs can be sufficient to establish in some cases that hundreds of thousands of model runs may be insufficient.

The sample sizes of 21 outer and 30 inner (demanding a total of 620) model runs are intended to be the smallest sample sizes that will give the correct order of magnitude for the bias. Note that the same 620 runs can be used when considering any inner sample size n.

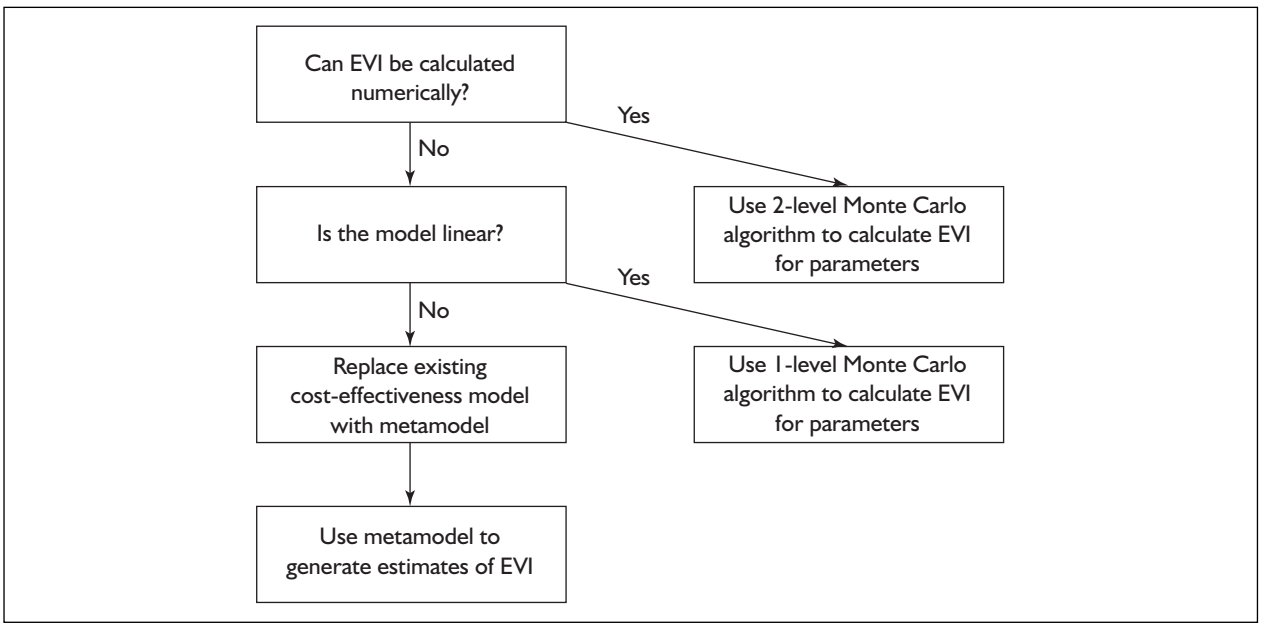

FIGURE 4 *Outline flowchart for implementing EVPI analysis*

BOX 5 *Method for calculating necessary sample size for EVPI analysis*

Sample size calculation
Let
θ_i be the parameters of interest for partial EVPI
$\theta_{\sim i}$ be the other parameters
Then:

1. Choose 21 evenly spaced values of θ_i across the parameter's sample space for the outer samples.
2. For each of the 21 values, sample $s = 30$ inner values of $\theta_{\sim i}$. Evaluate the mean net benefit of each treatment and the sample variance-covariance matrix for the d net benefits. Denote the mean vector and variance covariance matrix by M and V, respectively.
3. To estimate the bias using an inner sample size of n, generate a large sample of multivariate normal random vectors Z from the $N(M,V/n)$ distribution. Evaluate the average value of $\max\{Z\} - \max\{M\}$.
4. Repeat steps 2 and 3 for each of the 21 θ_i values. This will produce 21 biases calculated in step 3. Use numerical integration to obtain the expected bias with respect to θ_i.

Is the model linear?

Under some conditions, a one-level Monte Carlo integration can be used to evaluate the partial EVPI for individual and groups of parameters. The equations for calculating the EVPI for a parameter or group of parameters θ_i is presented in *Box 2* and set out as follows:

Let,

θ_i be the parameters of interest for partial EVPI
θ_{-i} be the other parameters
d be the set of possible decisions or strategies.

Partial EVPI obtaining data on θ_i only =
$$E_{\theta_i}\{\max_d(E_{\theta_{-i}}\,\mathrm{NB}(d,\theta)\,|\,\theta_i)] - \max_d\{E_\theta\,\mathrm{NB}(d,\theta)\}$$

The need for a two-level Monte Carlo integration arises from the double expectation contained within the first element of this formulation. However, the expected value of a vector function $Fn(\theta_{-i})$ is equal to $Fn(\overline{\theta_{-i}})$ if the function is linear [for example, $Fn(\theta_{-i}) = A_1 \times \theta_{-i(1)} + A_2 \times \theta_{-i(2)} + A_3 \times \theta_{-i(3)} + \ldots +$ constant], this is a sufficient but not necessary condition. The conditional nature of the expectations in the above expression also imposes a further restriction in that for all possible values of θ_i, the conditional expectation of $Fn(\theta_{-i})$ given θ_{-i} equals the unconditional expectation $Fn(\overline{\theta_{-i}})$. This therefore imposes a further condition, which is satisfied if θ_i and θ_{-i} are independent, but not generally otherwise. Where these conditions hold the first element of the above EVPI expression reduces to a single expectation:

$$E_{\theta_i}\{\max_d(E_{\theta_{-i}}NB(d,\theta)\,|\,\theta_i)\} \equiv E_{\theta_i}\{\max_d NB(d,\theta_i\,|\,\theta_{-i} = \overline{\theta_{-i}})\}$$

This simpler expression can be evaluated using a one-level Monte Carlo integration over the uncertainty in θ_i, where the inner bracket on the left-hand side, $(E_{\theta_{-i}}\,NB(d,\theta)\,|\,\theta_i)$, is replaced with the net benefit obtained when θ_{-i} are fixed at their prior means on the right-hand side.

The linearity of the model can be checked using standard statistical techniques applied to a sample of the overall model outputs obtained from a Monte Carlo analysis. Formally, for each parameter or group of parameters of interest, the linearity of the model in the parameters not of interest should be checked; however, if applied thoroughly the savings in sampling time aimed for would disappear. In practice, a single check of overall model linearity can provide a minimum investigation, although it should be noted that it is possible for the whole model to be non-linear, whilst remaining linear in the parameter subsets sufficient to satisfy the above conditions.

The method for calculating the partial EVPIs for individual or groups of parameters, using the one-level approximation, are the same as those for calculating the overall EVPI, with the exception that the random sampling in step 1 is undertaken only for those parameters of interest; the remaining parameters not of interest should be held constant at their mean values. The complete description of steps is set out in the section 'Calculating partial EVPI for individual or groups of parameters' (p. 5).

Metamodelling

Review of metamodelling techniques

Introduction to metamodelling

Computer simulation models are used because of the impracticalities involved in creating several versions of a real system, or because of constraints which prevent experimentation with a system. Simulation models, despite being simplifications of

the real system, can be extremely complex and their processing time for repeated simulation runs can impose severe restrictions on the amount of model analysis undertaken. Improvements in computer power and speed have not provided a solution to this problem, as practitioners seek to carry out simulation analysis in large-dimensional problems.

For this reason, one approach to modelling the original problem which has become more popular over the past quarter of a century is the use of metamodels as replacements for the original simulation models. Metamodels are effectively 'models of models',[47] or mathematical approximations to the input and output functions of a model,[48] and can be seen as simplifications of the original model. They are intended to provide a good approximation of the model while significantly cutting down the necessary computing time by reducing the processing time required, without oversimplifying the model. Metamodels have a number of alternative uses, including enhanced exploration and interpretation of the model, generalisation to models of a similar type, sensitivity analysis, optimisation and providing the analyst with a better understanding of the overall system.

Suppose that, for a given simulation model, the relationship between the input parameters and the output is given by the following (this can be generalised to problems involving more than a single response variable):[49]

$$z = f(\mathbf{x})$$

where $\mathbf{x}$ represents a vector of input parameters and z is the output. Metamodelling attempts to approximate the function f that relates the vector $\mathbf{x}$ with the output z. This process involves defining a separate function ϕ, with a predicted output $\phi(\mathbf{x})$, where ϕ must approximate f with sufficient precision, i.e. $f(\mathbf{x}) \cong \phi(\mathbf{x})$.

There are three main issues to consider in metamodelling, defined as:[50]

1. The choice of a suitable functional form for f (to represent the relationship between the inputs and the output).
2. Design of experiments [i.e. a definition of the input space over which the metamodel must approximate the response variable, the number of runs to simulate and the assignment of random number streams. See the section 'Data for metamodel building' (p. 20) for a discussion of the design of experiments].
3. Assessment of the adequacy of the model (using, for example, hypothesis testing and lack of fit analysis).

The first point addresses the choice of metamodel to be used. The second is concerned with how the metamodel is actually fitted (i.e. the range of values at which to observe y, and which parameters to use);[47] this involves adopting an appropriate strategy for selecting points in the design space for fitting the metamodel. The third point reflects on how well the metamodel approximates the original model.

Review of metamodelling: methods

We conducted systematic searches to identify studies relating to metamodelling techniques, in particular those studies that presented a comparative evaluation of alternative techniques. Systematic searches were conducted on the following databases: Computer and Information Systems Abstracts, the Conference Papers Index (via Cambridge Scientific Abstracts), International Abstracts in Operational Research (IAOR), Social Science Citation Index and Science Citation Index (via the Web of Science). The search terms used were broad, including either "metamodel" or "metamodelling" AND (compar* OR evaluat*). Internet search engines such as Google were also used to identify other literature that is not yet published in peer-reviewed journals. There were no restrictions on the basis of language, date or publication type. All searches were conducted between July and August 2003.

A number of different methods are available for the construction of metamodels; these have been classified into two categories: parametric and non-parametric techniques.[49] Parametric techniques approximate the functions with no prior knowledge about the underlying data; such techniques include polynomial models and Taguchi models. Non-parametric techniques have an *a priori* set of functions which are used to derive an approximate function based on observed responses. There is no single metamodelling technique that is universally applicable to all simulation models, and numerous factors affect the choice of a metamodelling approach, such as accuracy, efficiency, robustness, simplicity and model transparency.[51] Kleijnen and Sargent[52] recommend the following steps in building a simulation metamodel:

1. Determine the goal of the metamodel.
2. Identify the inputs and their characteristics.
3. Specify the domain of applicability.
4. Identify the output variable and its characteristics.

5. Specify the accuracy required of the metamodel.
6. Specify the metamodel's validity measures and their required values.
7. Specify the metamodel and review this specification.
8. Specify a design including tactical issues and review the design of experiments.
9. Fit the metamodel.
10. Determine the validity of the fitted metamodel.

Data for metamodel building

In order to construct a metamodel, data must be made available from which the relationships between the inputs and outputs of the simulation model can be estimated. This data collection process is known as design of experiments (DOE), and in a metamodelling context refers to the selection of the input variable settings to be sampled to build the metamodel;[53] it is therefore key to the development of a successful metamodel.[54] The use of an appropriate design is therefore critical to the performance of a metamodel, because the use of variable settings which are unrepresentative of the true model parameters will lead to the development of a metamodel which cannot accurately predict outcomes from new data. There are a number of criteria which can be used to measure an experimental design's capability, including:[53]

- the number of experimental runs required (of the simulation model)
- the symmetry of resulting distribution of variance around the design space
- the ease with which the design can be implemented
- the estimation of capacity of the design
- its ability to screen important factors.

There exist three key designs for fitting metamodels:

1. Full factorial design: this is the most commonly used design, because of its ability to sample from all dimensions of the design space via a series of uniformly spaced values. Each parameter is assigned a number of levels. As the number of parameters in the model increases, the number of design points required explodes; this can be solved by using fractional factorial designs. These designs assume that the higher order interactions are negligible,[55] enabling lower order effects to be estimated with considerably fewer runs.
2. Latin hypercube design: this design offers flexible sampling sizes whilst distributing points randomly over the design space.[54] Latin hypercube sampling involves dividing the distribution of each input parameter up into strata of equal probability and sampling once from each stratum to generate a set of different parameter values. This process can be repeated to generate multiple parameter sets. However, because the points are selected at random, it is possible to generate poor designs.[54]
3. Full factorial Latin hypercube designs: this is a hybrid design strategy which combines the above two strategies. The design space is divided into p^b hypercubes, with $p = k - m$, where k is the number of parameters and m is the fractionation of the design. In each hypercube, n points are generated using Latin hypercube sampling.[54]

The selection of an appropriate design of experiments for each metamodelling technique is discussed in the following sections.

Metamodel validation

An intrinsic objective of metamodelling is to derive a metamodel which has the capability of predicting the response over the design space of interest.[51] Hence, before being used as a prediction tool in place of the simulation model, it is important that any metamodel is validated to ensure its adequacy in prediction. Although it would seem convenient to assess the goodness-of-fit of a metamodel using the data used to build the metamodel (training data), this is considered an insufficient validation method.[51] Rather, additional samples should be taken from the simulation model, the results of which can be compared with the corresponding metamodel outputs. The availability of the original model means that a large volume of data can be generated to build the metamodel, which could be expected to improve the predictive power of the metamodel and can also be used to validate the metamodel (see below). The fit of the metamodel can be improved by deriving more samples from the original model to provide input into the metamodel, allowing it to be strengthened in areas of weakness; this is particularly important in ensuring that the metamodel provides a good fit in the areas in which it is known to attain most value. The accuracy of a metamodel can be assessed using three statistics:[51]

1. r^2 (a large value of which indicates an accurate model).
2. Relative absolute error (a small value of which indicates an accurate metamodel – often highly correlated with r^2).

3. Relative maximum absolute error (a small value of which indicates an accurate metamodel – generally not as useful as r^2 or relative absolute error).

In addition to these measures, it would be possible to calculate a confidence interval for the ENB of each treatment for the metamodel, which could be compared with that from the original model. Although this would give a rough gauge of the precision of the metamodel, it is not clear how these confidence intervals could be translated to provide a confidence interval of the EVPI.

Residual analysis of the fitted metamodel would enable the validity of the model to be assessed and highlight any potential problems which could affect this. An inadequate model can often be detected by simple plots showing patterns in the errors. For example, a plot of the predicted outcome values from the metamodel against the errors of these values would enable the analyst to determine whether the errors were both sufficiently small and random to give confidence in the metamodel's validity. Inadequacy of the model determined via residual analysis can often be addressed by the introduction of additional variables (for example, non-linear functions of the original variables) or by the transformation of variables. Residual analysis is particularly useful for the validation of regression models.

Metamodelling techniques

The review identified six metamodelling techniques. A brief introduction to each technique is provided and a critical assessment of their suitability to different simulation scenarios, and specifically their suitability for use in EVI analysis, is presented. The metamodels discussed are:

- linear regression
- neural networks
- multivariate adaptive regression splines (MARS)
- response surface methodology (using polynomial regression)
- Gaussian processes/Kriging (non-linear regression).

Simple linear regression

Simple linear regression explores the association between a dependent variable (y) and an independent variable (x). Regression analysis considers how variability in one or more independent variables causes a change in the behaviour of a dependent variable. For simple linear regression, where one is exploring the relationship between a dependent variable and a single independent variable, the relationship between the variables is summarised by a regression equation, which consists of a slope, an intercept and an error term. The slope represents the amount by which the dependent variable increases with unit increase in the independent variable. For simple linear regression, the relationship between the two variables is of the form:

$$y_i = \beta_0 + \beta_x + \epsilon_i$$

where β_0 is the intercept and β is the regression coefficient (slope). Therefore, over the range of values sampled from the population, a unit increase in x is expected to result in a change in y of β units. A least-squares approach is used to produce a model which minimises the prediction error between the model and the observed values from which the model is produced.

The strength of this relationship may be explored generally through the interpretation of a scatter diagram. The linearity of the relationship may be more accurately estimated using Pearson's product moment correlation coefficient, which is given by

$$r = \frac{\Sigma xy - \dfrac{\Sigma x \Sigma y}{n}}{\sqrt{\dfrac{\Sigma x^2 - (\Sigma x)^2}{n} \dfrac{\Sigma y^2 - (\Sigma y)^2}{n}}}$$

where $-1 \leq r \leq 1$.

The square of the r statistic essentially denotes the relative predictive power of a linear model, ranging between 0 and 1, and indicates how much of the variability in the dependent variable can be explained by the variability in the independent variable. For a strong positive correlation, the value of r will tend towards 1, whereas for a strong negative correlation, r tends toward –1. The adjusted r^2 statistic weights r for the number of independent variables and observations and hence penalises an unnecessarily complex linear model.

In reality, it is unlikely that the variability in the dependent variable will be completely described by a single independent variable. Multiple linear regression enables the variability in the dependent variable to be described by more than one independent variable. Whereas simple linear regression considers the linear relationship between two variables over two dimensions, multiple linear regression considers the linear relationship between a dependent variable and k independent variables over $k + 1$ dimensions. The

regression model is a simple extension of the simple linear form outlined above:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}$$

where β_0 is the intercept and x_{1i} is the first independent variable, x_2 is the second, and so on up to the kth independent variable.

Assumptions of linear regression

Linear regression is based on a number of assumptions:

- Each error component follows an approximately normal distribution, with a mean of zero. This can be evaluated by plotting the residuals, which should be randomly distributed about zero.
- Homoscedasticity of the errors – the variance of the errors in the predicted y values should be constant. This can also be validated by examining the residual plots and checking that there is no pattern in the residuals that suggests heteroscedasticity.
- The errors are independent of each other – a residual plot should show random scatter of the errors, and would highlight any patterns suggesting non-independence, such as autocorrelation.

Use of linear regression in EVI analysis

If a strong linear relationship exists between sample parameter inputs and the model outputs, it may be appropriate to use multiple linear regression to replace the existing cost-effectiveness model with a single regression equation. For the purposes of EVPI analysis, the dependent variable is defined as the absolute ENB for each treatment strategy and the independent variables are defined as the parameter values randomly sampled using multivariate Monte Carlo sensitivity analysis. The equation would include the many universal parameters common to all treatment strategies, plus strategy-specific parameters (e.g. treatment effect) which would be weighted according to the strategy under consideration. It should be noted, however, that there exists no formal test to answer the question 'when does linear become too non-linear'?; whilst the adjusted r^2 statistic is useful, the point at which a model becomes too non-linear is essentially dependent on the subjective judgement of the analyst and the level of accuracy required in the EVPI results. The relationship between the EVPI results of regression metamodelling and the original cost-effectiveness model should be highlighted as an issue for further research.

If the relationship between independent and dependent variables is weak, it may be necessary to specify a more complex model, for example, the inclusion of first-order interactions to obtain a better approximation. This would however increase the complexity of the metamodel; to specify a regression model with x independent variables would require $(x^2 - x)/2 + 2x$ variables if all first order interaction terms are included. Where a large number of variables are included in the model, this could make the metamodelling process as complex as implementing the original model itself. It may be possible, however, to specify a lesser number of interactions individually, rather than including the entire set. For the practical implementation of EVPI, regression analysis is likely to be most useful in instances where a strong linear relationship exists without having to include any interaction terms.

Does it save time?

Comprehensive value of information analysis may be performed using the regression model in considerably less time than is required for the equivalent analysis using the original cost-effectiveness model. The inclusion of first-order interactions may, however, require a more complex model.

Accuracy of results

The main drawback concerns the degree of linearity between the model inputs and the ENB. If the relationship between ENB and the parameter inputs is weak, linear regression is unlikely to be useful in performing EVPI analysis. If the relationship is strongly linear, that is, an adjusted r^2 value of close to 1, it is likely that even if the expected net benefits for each treatment strategy are predicted with accuracy, the prediction error in the calculation of net benefits is likely to be magnified in the calculation of EVPI. Despite this drawback, regression analysis may have an alternative role in identifying the key parameters within the model. Even if the predictions from the regression model are imperfect, the regression metamodel may be used in order to obtain one-level estimates of partial EVPI for all model parameters. Although the estimates of partial EVPI obtained through this method are unlikely to be accurate, the exercise may enable the modeller to ascertain which of the individual model parameters are likely to attain value and which are not. The fundamental benefit of this deductive approach is that if the analyst is aware of the key parameters, it may be possible to revert back to the original cost-effectiveness model and

perform partial EVPI analysis on those identified parameters, and to ignore the remaining parameter set.

The appropriateness of linear regression metamodelling is determined by the degree of linearity between the model inputs and ENBs. The question of 'how linear is linear enough?' for use in EVPI analysis cannot be resolved using standard statistical tests, although it is possible to explore the degree of approximation error resulting from a linear regression metamodel through comparing the global EVPI results calculated using the regression metamodel and the global EVPI results calculated using the original cost-effectiveness model. If the two global EVPI results are similar (and the adjusted r^2 value suggests a strong linear relationship), this should enable the analyst to gauge the degree to which non-linearity may distort the results of the partial EVPI analysis. If there is a considerable error between the global EVPI estimates, this should forewarn against the use of the one-level EVPI algorithm and highlight the need for non-linear methods such as GP metamodelling.

Level of complexity/specialist expertise required
One of the key benefits of using regression analysis in undertaking value of information analysis is that the method is straightforward to implement and a regression metamodel may be constructed in only a short amount of time.

Are there any parametric restrictions?
There are no formal restrictions in terms of the number of parameters that may be used within the regression analysis. However, if a large number of parameters are included within the regression model, a greater number of data points are required in order to cover the entire response surface. It is generally recommended that 10 sets of observations are required for each parameter within the regression model, although it may be beneficial to use a greater number of observations to increase the accuracy of the approximation. The inclusion of first-order interactions may be constrained by the ability of the software to cope with a large number of variables and observations.

Neural networks

Neural networks are a class of non-parametric models capable of learning from data. Originally devised in the 1960s, the idea of developing simplified mathematical models of brain-like systems did not become widely used until the late 1980s, but they are now used in areas such as classification (for example, in marketing, as a means of targeting customers by classifying them into segments based on certain characteristics), financial analysis and optimisation.[48] A neural network is intended to mimic the functioning of the brain by 'learning' about the system being modelled based on data fed into it. These data often come from simulation model runs, and are known as the 'training' data, which comprises values of the input parameters and the corresponding outputs. The network uses this information to learn about the relationships between the various inputs and outputs, based on different values of the input parameters. This knowledge about the simulation model then allows the network to make estimates of the outcome(s) based on new sets of input values, saving computation time by removing the need to run many simulations with different parameter values. A neural network is made up of a series of nodes, which are arranged in layers. Data on the simulation model parameters are fed into the network and propagated through the network via these nodes to produce an output value.[48] The layout of a typical node is shown in *Figure 5*.

Each node accepts one or more inputs (e.g. x_1, x_2, x_3) and subsequently produces an output based on weights (e.g. w_1, w_2, w_3) assigned to each input to represent the strength of the relationships between the nodes.[48] Each node combines its inputs and adds a bias term (θ) to give a net activation value. The output of each node (y) is then a non-linear (often logarithmic)[56] function of the net activation value. This output may form an input to another node, or it may be the final output of the network.

There are a number of ways in which the nodes can be arranged in a neural network (known as **architectures**), the most common being the feed-forward network, in which the nodes are arranged in layers where the output of each node is connected to all of the nodes in the next layer.[48] Under this system, the procedure described above continues until the output layer is reached.

In order to train a neural network, a suitable set of data needs to be fed into it; this usually comes in

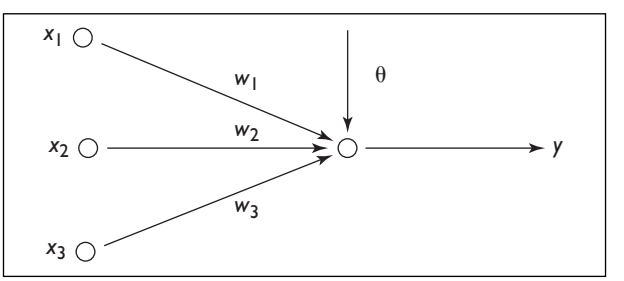

FIGURE 5 *Example node of a neural network*[56]

the form of results from the original simulation model. The training is carried out by repeatedly giving the network examples from the training data set.[56] Each component of the training data consists of an input vector (i.e. set of input parameter values) along with its corresponding response (output). Each example is fed into the network, which calculates an estimated value of the output, and the resulting error between this estimate and the true answer is used to adjust the weights (w_i) between each layer in the network.[56] The most common method of learning is known as back-propagation,[56] where the errors are propagated through the network from the output layer back to the initial input layer. The quantity and quality of this training data are key to the performance of the network, and as such neural networks are sometimes limited in their prediction capability by finite and imperfect data sets.[57] Hence, the larger and more representative the training data is of the true system, the more accurate the network will be in predicting the output.

A trained neural network metamodel is a deterministic model,[57] but neural networks can also be used to model stochastic systems. However, this is hindered by the need to obtain each individual training point, which can involve vast computational expense. As a result, the training data sets used for stochastic networks tend to be small, and this in turn leads to a less precise network, yielding potentially inaccurate predictions.

The use of neural networks in metamodelling has a number of advantages, primarily that they can theoretically represent any relationship to any degree of precision because they have universal function approximation capability,[48] and at the same time can significantly accelerate the computer time required for large simulation models. As a result, the error in the network can theoretically be reduced to zero because the likelihood of selecting an incorrect functional form is eradicated.[57] Neural networks are global models, and so unlike in polynomial regression, a single network can model the entire simulation response surface.[57] They do not rely on some of the standard statistical assumptions such as homoscedasticity (constant error variance), absence of autocorrelation and absence of multicollinearity between the input variables; they can accommodate a combination of continuous variables and discrete numeric variables.[57] Since many simulation models have multiple outputs, the use of neural networks avoids the need to develop a separate metamodel for each output.

As with any metamodelling technique, the network is only valid over the specified parameter domains included in the training data set, which emphasises the need for a reasonably large training set. The time taken to train the network may be substantial and a large amount of training data often needs to be simulated from the original model to train the network adequately. The determination of a suitable architecture for the network can also be time consuming and, because of its nature as a non-parametric approach, a network may over-fit the random error obtained from the simulation.[48] If additional parameters are required in the model or the distribution of any input parameter is changed, the network is invalidated and a new network must be created to reflect these changes, which inevitably increases computation time. It is ultimately the decision of the analyst as to whether the effects of some of these problems are offset by the speed with which networks can compute answers, once trained.

Software is widely available for the development of neural network models, for example neural simulation tool (NEST),[58] giving it an advantage over other metamodelling techniques.

Use of neural networks in EVI analysis

Does it save time?

Neural networks are particularly useful in dealing with computationally expensive models; from this point of view, using neural networks may save time compared with a large simulation model. However, as the number of parameters increases, so too does the level of training required to calibrate the model. Problems may arise in training the network; because of the large number of parameters in the model, the training process may be lengthy because of the need to offer the network training data on all parameters at a variety of different parameter settings. The extent to which this is done is dependent on the level of accuracy required in the output values from the network compared with those from the simulation model. Therefore, the timescale of the project would determine the feasibility of using neural networks.

Accuracy of results

The accuracy of predictions made by a neural network depends again on how intensive the training process has been. If a sufficient amount of data which is representative of the simulation model has been fed into the network, its ability to predict the outcome based on new data will be improved. A rigorous training process is therefore recommended if highly accurate approximations are required.

Level of complexity/specialist expertise required
The transition from a deterministic health economic model to a neural network may not be straightforward and specialist expertise may be required. Although a suitable software package may ease the process of building a neural network metamodel, it is one of the more complex metamodelling methods identified within this review.

Are there any parametric restrictions?
One of the main advantages of this method is that it can theoretically handle any number of parameters, although this may be limited by the capacity of the software used. Over the lifetime of a long-term project, a neural network could effectively replace the simulation model,[56] with the original model being used as a validation tool against which to check the accuracy of the network's predictions.

Multivariate adaptive regression splines

In many situations, the behaviour of a simulation model cannot be adequately described by a single equation. If the data under consideration are highly non-linear, one approach would be to use as high-an-order polynomial as possible to approximate the function. However, high-order polynomials can be inappropriate in such situations because they rely on the cancellation of oscillations to obtain a good fit; this property makes them non-robust.[47]

This problem led to the development of spline metamodels, which address the problem of non-linearity by fitting a set of low order polynomials, or splines, to the data, each over a separate range. These ranges are derived by dividing the domain up into intervals, such as $[t_1, t_2]$, $[t_2, t_3]$, …, $[t_{n-1}, t_n]$; the end-points of these intervals are known as knots.[47] Continuity restrictions are applied to adjacent pieces to ensure that the pieces match with a prescribed order of continuity.[53] Each range is then represented by a different equation. The univariate spline metamodel can be represented as follows:[59]

$$\hat{y} = \sum_{m=1}^{M} a_m B_m(x)$$

where $\hat{y}$ is the outcome and a_m is the coefficient of the expansion (or weighting) applied to the basis function, B_m. These basis functions typically take one of two forms:[47]

- truncated power function bases
- B-spline bases.

The main issue in spline metamodelling is the trade-off between the fit of the approximation at known points and the smoothness of the resulting metamodel.[47] The fit is measured by the sum of squared differences of the metamodel and the responses of the simulation model in each experimental run, and smoothness is represented by integrating the square of some derivative over the region of the metamodel's validity. The relative importance of these two objectives is defined by the smoothing parameter λ.[47] For example, $\lambda = 0$ implies interpolation with no constraint on smoothness. The function which minimises this value is a spline of order k, which is in C^{k-2} [continuous derivatives up to the $(k - 2)$th derivative].[47] This function is a piecewise polynomial with terms up to x^{k-1}, and the knots will occur at points in x corresponding to the observed data, x_j.

There are three classes of spline metamodelling methods which are used to deal with the trade-off outlined above:[47]

- smoothing splines
- spline interpolation
- regression splines.

Under the smoothing splines approach, the order of the spline, k, is chosen by the user; knots are not pre-specified, but they will occur at the x_j values in the optimal solution; λ can be chosen based on the generalised cross-validation (an adjusted residual sum of squares). Spline interpolation also requires k to be chosen by the user; knots are not pre-specified, but they will occur at the x_j values in the optimal solution, while $\lambda = 0$. Finally, under the regression splines methodology, k and the knots are chosen by the user, and $\lambda = 0$.[47]

Such univariate spline metamodels can be extended to multivariate scenarios; the most common approach in these situations is the use of MARS. This method works by adaptively selecting a set of basis functions for approximating the response function through a forward/backward iterative approach.[51] The method is implemented as in the following steps:

1. Start with the simplest model containing only the constant basis function.
2. Search the space of basis functions, for each parameter and for all knots, and add those basis functions which maximise the goodness-of-fit.
3. Recursively apply step 2 until a specified level of complexity is attained.

4. Remove the basis functions which contribute the least to the overall (least-squares) goodness-of-fit.

This method makes no assumptions about the input–output relationship and has therefore become increasingly popular in dealing with complex data mining problems because it can be used in situations where the relationship between the inputs and outputs is non-monotonic and hence difficult to approximate with parametric techniques.[59] In addition, step 4 makes MARS a very powerful tool for parameter selection, as the algorithm only picks up basis functions and predictor variables which make a sizeable contribution to the prediction of the outcome.[59]

The main advantages of MARS are its accuracy and the reduction in computation time required to construct the metamodel compared with other metamodelling techniques.[51] In addition, it is able to deal with both continuous and discrete variables (i.e. the input parameters). However, because the method is adaptive, it offers a high degree of flexibility, which may result in overfitting.[59] Models which are unnecessarily complex often tend to be poorly generalisable to the prediction of new cases (i.e. perform poorly when presented with new data).[59] This problem is overcome in part by step 4, but does not necessarily guarantee to reduce complexity sufficiently. Spline metamodels have also been accused of being difficult to interpret.[48]

Software is available for building MARS metamodelling, the most widely used being Matlab because of its flexibility for creating algorithms.[53]

Use of multivariate adaptive regression splines in EVI analysis

Does it save time?

In order to determine suitable basis functions to construct the metamodel, a sufficiently large training data set would need to be produced via multivariate Monte Carlo sampling in order to reflect the varying levels of uncertainty in the input parameters. Non-linearities may occur when the MARS approach is used for models with a large number of parameters, thus making parametric metamodelling techniques inappropriate. MARS does not make any such assumptions about the linearity of the model and, because of its use of multiple splines to model the data, can incorporate varying degrees of non-linearity within the same metamodel. Since the aim of a metamodel is to simplify the original simulation model, the property of removing (one by one) the least significant basis function from the metamodel eliminates the need for initial factor screening, saving computation time and the need to select an appropriate screening technique [see the section 'Importance analysis' (p. 31)]. MARS requires less computation time in constructing the metamodel than many other techniques, hence it may be a useful technique if time is limited.

Accuracy of results

The use of MARS is thought to yield more accurate results than low-order polynomial approximations, because of its use of piecewise polynomials as opposed to a single polynomial equation. The ability of MARS to predict responses accurately is highly beneficial in value of information analysis as the magnitude of errors made in the calculation of the net benefits is augmented in the EVPI calculation.

Level of complexity/specialist expertise required

Although the computation expense involved in constructing MARS metamodels is generally low, it is a relatively complex method, requiring some statistical expertise in deriving the basis functions from the simulated data. This complexity can lead to overfitting, yielding an unnecessarily complex model which is difficult to interpret.

Are there any parametric restrictions?

There exists no specific limit on the number of parameters which can be used in a MARS metamodel, but because of the computation required to derive the basis functions, a low level of dimensionality not only reduces the computation time required, but also reduces the possibility of overfitting the model.

Response surface methodology

Response surface methodology is described as "a collection of statistical and mathematical techniques useful for developing, improving, and optimising processes. It consists of the experimental strategy for exploring the space of the process or of independent variables, empirical statistical modelling to develop an appropriate approximating relationship between the response and the input variables."[60] It has been used as a technique for metamodelling for over 30 years, and works by fitting a series of polynomial regression models to the output variable of the simulation model and optimising the resulting regression function.[61]

The metamodel takes the form:[62]

$$y(x) = f(x) + \epsilon$$

where $y(x)$ is the unknown function of interest, $f(x)$ is a known polynomial and ϵ is a random error following a normal distribution with mean 0 and variance σ^2. The polynomial function $f(x)$ which is used to approximate $y(x)$ is typically a first- or second-order polynomial (i.e. it is either linear or quadratic). For example, a first-order polynomial would take the form:

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i$$

where β_0 is the intercept term, x_i is the input parameter i and β_i is the coefficient of parameter x_i.

Metamodels based on polynomial regression tend to be popular because they provide parameter estimates which have advantageous properties such as being unbiased. The parameters of these polynomials (β_i) are determined using least-squares regression analysis of the sampled data from the original simulation model as follows:[62]

$$\beta_i = [\mathbf{x}'\mathbf{x}]^{-1}\,\mathbf{x}'\mathbf{y}$$

where $\mathbf{x}$ = the design matrix of data points sampled from the simulation model, $\mathbf{x}'$ = the transpose of $\mathbf{x}$ and $\mathbf{y}$ = a column vector containing the values of the response at each sample point.

These coefficients are used to fit the response surface approximations to data generated from the original simulation model, which are then used for prediction of the outcome variable of interest.[54]

The process of response surface metamodelling encompasses three steps:[62]

1. Screening of parameters in the simulation model to select the most important ones (if the model contains a large number of parameters) – see the section 'Importance analysis' (p. 31) for details of screening methods.
2. First-order (polynomial) experimentation – this considers a linear method of approximating the input–output relationship of the simulation model if little curvature exists in the data.
3. Second-order (polynomial) experimentation – this follows on from first-order experimentation if significant curvature exists.

The use of low-order polynomials is advantageous because they involve relatively few parameters and offer insight into the model behaviour. However, because most response surfaces are second-order designs, they have limited capability in accurately modelling non-linear functions of arbitrary shape.[62] Although higher order response surfaces could be used to overcome this problem, this creates more problems because it requires a larger number of simulation runs to be carried out to provide the data points necessary for fitting models with a large number of variables. This can often involve considerable computational expense. The degree of linearity in the model would need to be estimated in advance (for example, using the r^2 statistic) to enable an appropriate polynomial approximation to be used.

A potential solution to this problem which has been suggested is the use of multi-level, partitioned response surfaces, in which parameters and responses are divided into two groups based on domain knowledge, that is, based on factors which are believed to affect particular responses directly.[63] Each response is made a function of all other factors by concurrently constructing two-part partitioned response surfaces.[63] This involves fitting the first set of responses as a function of the first set of factors and fitting the second group of responses as a function of the second set of factors. Separate experiments are then run to fit these two sets of response surfaces. The effect of the second group of factors on the first set of responses is assessed by fitting the mean of these responses as a function of the second set of factors and vice versa, creating two-level response surfaces.

This method is based on the assumption that the interactions between the parameters of each partitioned set are either negligible or non-existent. This can be overcome in part by ensuring that the factors and responses are partitioned appropriately, which should minimise the magnitude of these interaction terms.

Use of response surface methodology in EVI analysis

Does it save time?

The ability of the response surface methodology in reducing computation time depends on the level of non-linearity in the simulation model being approximated. If a low-order polynomial provides an adequate approximation (i.e. if the model is approximately linear), then the computation time involved in deriving the metamodel is negligible. If, however, the data are highly non-linear and the use of a higher order polynomial is required, this can increase the computation time considerably

because of the need to simulate additional data points to build the metamodel.

Accuracy of results
The accuracy of the results provided by a response surface metamodel depends on the order of the polynomial used. This decision is driven by the level of non-linearity present in the data, and although a low-order approximation may prove sufficient in some cases, data with a significant non-linear pattern may be better approximated by a higher order approximation. As with many metamodels, there exists a trade-off between the accuracy required from the metamodel and the time available in which to build it.

Level of complexity/specialist expertise required
The use of low-order polynomials in response surface metamodelling ensures that the models derived are relatively simple in terms of both their construction and interpretation. Specialist statistical knowledge may be required if higher order polynomials are to be used.

Are there any parametric restrictions?
The use of multi-level partitioned response surfaces may be appropriate in EVI analysis, because of their ability to deal with a large number of parameters. This method can be extended to use more than two sets of factors and responses and, given that parameters within many health economic models can often be grouped relatively easily (e.g. cost parameters, utility parameters, drop-out parameters), this approach could be beneficial. In addition, the experimentation and model fitting expense is greatly reduced by splitting up the responses and factors in this way.[63]

Gaussian processes

GP regression[64] is a Bayesian non-parametric method and has become widely used recently in approximating deterministic simulation models.[65] Operationally it is equivalent to Kriging, a technique used in geostatistics for spatial interpolation. The basic GP model is designed for interpolating deterministic functions, but can be modified to incorporate noise in simulation models such as that arising in a patient simulation model. GPs are particularly appropriate in cases where the simulation model under consideration is highly non-linear and there is uncertainty regarding the true functional form of the model. With sufficient data a GP can replicate any continuous deterministic function precisely.

In GP regression, the simulation model $y(x)$ is thought of as an **unknown** function, and uncertainty about this function is described statistically. *A priori*, it is stated that

$$y(x) = f(x) + Z(x)$$

where $f(x)$ is a known polynomial function of x (similar to the polynomial function in a response surface model); coefficients in the polynomial are treated as uncertain; and $Z(x)$ is the uncertain functional departure from that polynomial, represented by a Gaussian process.

Defining $Z(\cdot)$ to be a Gaussian process means that for any collection of inputs $\{x_1,\ldots, x_n\}$, the corresponding values of $\{Z(x_1),\ldots, Z(x_n)\}$ have a multivariate normal distribution. The implication of this is that at any two input values x_i and x_j, the functional departures $Z(x_i)$ and $Z(x_j)$ from the polynomial $f(\cdot)$ will be correlated, the size of the correlation depending on the distance between x_i and x_j.

The GP will have mean 0 [a non-zero mean function would be incorporated into $f(\cdot)$] and covariance:

$$\mathrm{Cov}[Z(x_i),Z(x_j)] = \sigma^2 R(x_i,x_j)$$

where σ^2 is the variance and describes how far $y(x)$ might deviate from $f(x)$ and $R(x_i,x_j)$ is the spatial correlation function between two sampled data points $Z(x_i)$ and $Z(x_j)$.

A variety of possible correlation functions exist, the most frequently used of which is

$$R(x_i,x_j) = \mathrm{e}^{-\theta|x_i - x_j|^p}$$

where $\theta > 0$ (estimated by maximum likelihood) and $p \leq 2$ [$p = 2$ is commonly used, and implies that $y(\cdot)$ is infinitely differentiable].

Functions of this nature tend to zero as $|x_i - x_j|$ increases. This indicates that the influence of a data point on the point to be predicted becomes weaker as the two points become more distant from one another, whereas the magnitude of θ determines how quickly that influence deteriorates.[53] The selection of which correlation function to use determines how the model fits the data;[53] it is dependent upon what level of smoothing between the sampled points is required and how quickly the function is required to move between them.

The construction of the GP metamodel, however, can be time consuming, owing to its complexity relative to other methods (e.g. response surface metamodelling). The difficulty arises in estimating

θ, as evaluating the likelihood can itself be computationally intensive.

Use of Gaussian processes in EVI analysis

Like the MARS approach, the use of GPs would be most beneficial in a situation where the model being approximated was highly non-linear. GPs have the additional advantage that the theory can be naturally extended to efficiently approximate integrals of functions $y(x)$ in addition to functions themselves. Since evaluating a partial EVPI involves evaluating a large number of integrals (expectations), the GP approach can be particularly efficient.[66] Difficulties are likely to be encountered when the number of input parameters is large (say >50). This is because the more inputs there are, the more model runs will be needed for the GP $Z(\cdot)$ to 'learn' how $y(\cdot)$ deviates from $f(\cdot)$. This poses a problem in that many health economic models are complex and contain a large number of parameters; this can be resolved by using importance analysis techniques [see the section 'Importance analysis' (p. 31)] to rank the parameters according to their importance in the model.

Does it save time?

The greater the complexity of the health economic model (and the time needed for a single model run), the greater the time that is saved by using a GP metamodel. Alternative metamodelling techniques may prove simpler and more accessible to the analyst, but will require more runs of the economic model before they can be constructed. This approach is not recommended for models that can effectively be run instantaneously.

Accuracy of results

Because the GP model is a Bayesian technique, it is possible in principle to quantify probabilistically the uncertainty in the true value of a partial EVPI that results from using a GP to emulate the model based on a limited number of model runs. The issue of uncertainty due to the use of a metamodel has been addressed,[66] but the methods have not as yet been adapted for partial EVPIs. Partial EVPI estimates obtained from a GP metamodel based on 600 simulation model runs were more accurate than Monte Carlo estimates based on 400,000 runs.[66] The simulation model in this example had 23 uncertain input parameters.

Level of complexity/specialist expertise required

GP regression is the most complex of the methods presented in this review, owing largely to the absence of suitable software available to construct this form of metamodel. The implication arising from this is that analysts must rely on their own statistical programming abilities. Until reliable software for implementing GP regression (and producing partial EVPI estimates) is readily available, this technique may be unattractive to a general user of health economic models.

Are there any parametric restrictions?

GP metamodels can be difficult to construct for simulation models with large (e.g. >50) numbers of parameters. Large numbers of input parameters can be dealt with, but this is computationally demanding and may require programming ability in a faster language such as C. As a result, importance analysis may be required for larger models; this may represent an important concern, as there is a chance that the importance analysis may fail to identify all important model parameters. Hence, the resulting GP metamodel may not be based on those parameters which have the greatest EVPI. Current GP methodology also requires the simulation model to be a continuous function.

Comparison of metamodelling techniques

There is very little literature which explicitly compares metamodelling techniques against each other. However, it has been contended that the various techniques should be compared only on the basis of accuracy of results.[51] However, for the general user of health economic models, other factors should be taken into account, such as robustness, efficiency, model transparency and simplicity.[51]

Seven criteria can be used for the comparison of metamodels:[53]

1. The ability to gain insight into the model (for example, the relationships between the input and output parameters).
2. The ability to capture the shape of arbitrary smooth functions based on observed values which may be perturbed by stochastic components with general distribution.
3. The ability to characterise the accuracy of the fit through confidence intervals.
4. The robustness of the prediction away from observed (x, y) pairs.
5. The ease of computation of the approximant function f.
6. The numerical stability of the computations and consequent robustness of predictions to small changes in the parameters defining f.
7. Whether software exists for computing the metamodel, characterising its fit, and using it for prediction.

Table 5 gives the main advantages and disadvantages of each metamodelling approach, along with brief discussion of its applicability to EVPI analysis.

There exist a number of additional metamodelling techniques, which are not presented here owing to a lack of relevant literature. These methods include:[67]

- accumulated approximation technique (which accumulates the values of previously used points)
- multivariate Hermite approximation (which requires a large number of data points)
- wavelet modelling (which uses a special form of basis function)
- project pursuit regression (which is useful for large models).

TABLE 5 *Advantages and disadvantages of each metamodelling technique in EVPI analysis*

Metamodelling technique	Advantages	Disadvantages	EVI application
Linear regression	Simple to perform Software widely available Useful if the simulation model is linear	Not useful if the model is non-linear Incorporating interaction terms into the model can make the analysis unnecessarily complicated	Estimate of EVPI may be distorted if the model is imperfectly linear
Neural networks	Very quick to run once trained Particularly useful for deterministic applications High accuracy can be obtained if network is sufficiently trained	Training process can be time consuming Modelling stochastic systems can yield imprecise results Software can limit the number of parameters which can be used Insufficient training can lead to a lack of robustness	May potentially be used to replace simulation model
MARS	Low computational expense in building the model More accurate than low-order polynomial approximations Flexible software available Ability to accurately capture shape of functions via the use of piecewise polynomials	The high degree of flexibility can result in overfitting Unnecessary complexity can make interpretation difficult Some specialist statistical expertise required for deriving the basis functions	MARS removes the need for factor screening
Response surface methodology	Simple to derive the metamodel Software widely available The use of low-order polynomials ensures simplicity	The adequacy of the model is determined solely by systematic bias in deterministic situations Using high-order polynomials increases computation time significantly	There may be a trade-off between accuracy required and time available
GP	Extremely flexible Well suited to deterministic applications Requires fewer parameters to be fit than other methods	Limited use for models with >50 parameters Complex method Lack of suitable software Requires specialist expertise in smoothing	EVI analysis may be problematic due to parameter restrictions and complexity

Few of these methods are currently widely used and they are the subject of ongoing research.

Conclusions

There is clearly no universally superior metamodelling technique, rather each method may be appropriate in different circumstances, depending on the size of the original model and whether the relationship between the input and output parameters is linear. Although metamodels allow faster analysis of a problem, their use introduces an added element of uncertainty to the analysis, because a metamodel can only ever approximate a system rather than fully replace it. By approximating net benefits in an EVI model, the accuracy of the subsequent calculation of the EVPI can be affected if the estimates of the net benefits are poor. The value of the EVPI is calculated from extreme values of the net benefit, so the approximation of the EVPI will be poor if the metamodel makes poor estimations at these extremes.

Because data from the simulation model are required to build the metamodel, the use of an appropriate design of experiments is critical to the success of the chosen metamodel. The use of data which are not representative of the true system in building the metamodel will result in bias and hence inaccurate predictions.

The most salient problem concerning certain forms of metamodelling (e.g. Gaussian process modelling) is that they are able to accommodate only a limited number of variables. The process of deriving a metamodel which accurately approximates the original model may therefore be problematic, owing to the necessity of having to ignore some parameters. There is hence a need to prioritise the parameters in some way so that the variables which most influence the outcome measure are used in the metamodel. There exists a variety of methods for prioritising parameters, which are discussed in the following section, along with suggestions as to how they could be applied to value of information analysis.

Importance analysis

Introduction

The main problem with some forms of metamodelling is their inability to accommodate a large number of variables. It is therefore often not feasible to use all input parameters; indeed, this is one of the reasons why metamodelling is used in the first place, as a means of simplifying the original model.

Importance analysis refers to a set of techniques which aim to rank parameters in a complex simulation model according to their relative importance in the model (i.e. their impact on the variability of the outcome), identifying the most important parameters and separating them from the trivial many. Such techniques provide the analyst with information about which parameters contribute most to the uncertainty in the outcome measure and hence to select these variables as inputs into metamodelling. It is therefore useful in situations where the metamodel being employed can handle only a limited number of variables. Importance analysis identifies the most influential variables for use in the metamodel, with the anticipated result being a good approximation of the original model.

In this respect, importance analysis is different from traditional sensitivity analysis, where the aim is to address the degree of uncertainty surrounding the outcome measure. It can also be used as a means of prioritising further research by providing information on which variables cause uncertainty and so should be researched in more depth.

There exist a number of importance analysis techniques for ranking parameters, which have been broadly classified into five categories:[68]

- variance/correlation-based measures
- information-based measures
- probability-based measures
- entropy-based measures
- elasticity-based measures.

Correlation- or variance-based methods tend to be the most commonly used because of their relative simplicity; such methods include assessing correlation and regression coefficients. Probability-based measures compare two cumulative density functions for a single-point parameter and measure the difference between them. Elasticity-based measures consider how the value of the outcome parameter is affected by changes in any one input parameter. Entropy methods measure the degree of dispersion of values for an outcome measure,[68] whereas information-based methods include the use of EVPI to measure the reduction in opportunity loss associated with eliminating all uncertainty on a parameter.

This section discusses a number of importance analysis techniques from the categories given

above, together with an assessment of how they may be applied to EVI analysis. The following techniques vary in their complexity and in their suitability for certain types of data; this is discussed when assessing each method's applicability to EVI analysis:

- generalised sensitivity analysis
- sequential bifurcation
- partial contribution to variance
- partial rank correlation coefficient
- standardised regression coefficient
- frequency domain methodology
- elasticity measures.

Generalised sensitivity analysis

Generalised sensitivity analysis (GSA) is a probability-based measure of which there has been relatively little discussion in the literature. This method considers the contribution of each model parameter to the eventual outcome measure and then uses the Kolmogorov–Smirnov (K–S) statistic to rank the parameters and prioritise further research by eliminating the less influential variables.

For each parameter of interest, two cumulative density functions (CDFs) are derived, relating to 'good' and 'bad' outcomes.[68] These two CDFs are plotted and compared by calculating the maximum separation distance (MSD) between the two curves, denoted by d. If there is a large separation distance between the two CDFs, then the response variable is sensitive to changes in that parameter. This measure of sensitivity can be assessed using the K–S test, which tests for a difference between two distributions. The K–S test also provides the probability of the MSD between the two CDFs that could have occurred if the two distributions in fact came from the same distribution.[69] Therefore, the lower this probability is, the more certainty there is that the CDFs are significantly different from each other. The K–S statistic is calculated for each parameter in the model; the parameters can then be ranked in descending order according to this statistic; parameters with the highest K–S statistic are those which cause most uncertainty in the outcome variable.

Use of generalised sensitivity analysis in EVI analysis

With respect to value of information analysis, this methodology could be adopted by plotting the CDFs of the ENB (with a positive net benefit representing a 'good' outcome and a negative net benefit a 'bad' outcome) for each parameter, comparing each treatment strategy with no treatment. By ranking the parameters based on the MSD or K–S statistic, the most important parameters could be determined and used to form the inputs to an appropriate metamodel.

Sequential bifurcation[70]

Sequential bifurcation (SB) is a group-screening technique which attempts to identify a set of n important parameters from a total set of N parameters in a simulation model. In some cases, N may be very large, which could make the process of identifying important parameters very time-consuming. SB is considered to be one of the better techniques in this instance because it requires fewer simulation runs than many other group-screening methods to identify the most important parameters, i.e. it is both effective and efficient.

The method makes a number of assumptions:[70]

1. The input/output behaviour of the simulation model can be approximated with a first-order polynomial (which looks only at the main effect of each input parameter on the output).
2. The errors from this polynomial are negligible.
3. The direction of the influence (i.e. positive or negative) which each parameter has on the output is known.
4. The total number of parameters, N, is a power of 2.

The method works by first transforming the model parameters into standardised variables, taking either the value 0 or 1; this enables the model to be approximated using a first-order polynomial. Each of these variables is assigned an upper and lower level which generate a high and low value for the outcome, y. These are defined as H_j (which generates a high value of y) and L_j (which generates a low value of y). These definitions, in conjunction with assumption 3 above, imply that the main effects of each factor are all non-negative. Therefore, all parameters with a main effect of zero are considered unimportant, whereas those which have a positive effect are important (the level at which a parameter is deemed important is discussed later). Now let the symbol $y_{(j)}$ denote the value of y (the simulation output) when factors 1 to j are switched on (take their higher levels, so the corresponding standardised variables take the value 1) and the remaining factors (N_{-j}) are switched off (take their lower values, so the corresponding variables take the value 0). The polynomial now takes the following form:[70]

$$y_{(j)} = \beta_0 + \beta_1 + \beta_2 + \ldots + \beta_\varphi \qquad \text{for } j = 0 \text{ to } N$$

where β_j represents the main effect (coefficient) of factor j.

Suppose that there are $N = 64$ parameters in the simulation model. A comparison is now made between $y_{(0)}$ and $y_{(64)}$, that is, setting all parameters at their lower level [$y_{(0)}$] and then all parameters at their higher level [$y_{(64)}$]. The expectation would be that $y_{(64)} > y_{(0)}$, implying that the sum of the main effects is important; if this is not the case, then none of the individual main effects is important. If this is the case, the group of factors is bifurcated (split into two groups of equal size). In the example given above, the initial split would separate factors 1–32 from 33–64 (if the total number of factors, N, is not a power of 2, then the factors are divided up by splitting them by the highest available power of 2; for example, if there were 100 factors, the first sub-group would contain 64 factors, with the remaining 36 in the second subgroup).

The sum of the main effects of the first 32 variables [$y_{(32)}$] would now be compared with $y_{(0)}$; if these values proved to be equal, this would indicate that none of the individual main effects of the first 32 factors was important. It would also compare $y_{(64)}$ and $y_{(32)}$; if these values were found to be different, it would indicate that at least one parameter from parameters 33–64 was important. This process would continue, splitting parameters 33–64 up into smaller and smaller groups and comparing the values of the resulting ys, until the most important parameters were identified. Because the process is sequential, the analyst need not specify in advance a critical value which a parameter's main effect must exceed in order to be considered important. Once the effect of a subgroup of variables is considered sufficiently small, the investigation of that subgroup can be stopped.[70]

The SB process can be extended to take into account the possibility of interactions between the factors; this is a relatively straightforward procedure,[70] but requires twice as many simulation runs to be made as for the individual main effects method. It should be noted, however, that the SB method is not appropriate when only interaction effects are important.

Use of sequential bifurcation in EVI analysis

The process of identifying important parameters via the use of SB is more time consuming than by some of the other techniques discussed, but as with these other methods, the most appropriate approach would be to use the ENB of an intervention as the outcome of interest. Each parameter within the EVI simulation model would be standardised to give a first-order polynomial approximation of the model, which would enable the main effect of each parameter on the ENB to be estimated (a second-order approximation could be used if it were suspected that interactions between parameters may be important). The upper and lower levels (H_j and L_j) would be determined using the analyst's knowledge of the model. For example, an increase in a parameter such as the cost of the intervention could be expected to reduce the expected net benefit and vice versa (in which case $L_j > H_j$), whereas an increase in the intervention's associated utility would increase the expected net benefit (hence $H_j > L_j$).

Given that SB is appropriate in situations where only a small number of parameters are important, it would be beneficial at this stage to attempt to group together these parameters and separate them from the less important ones. When the parameters are bifurcated into two equal-sized groups, the presence of all of the important parameters in one group would speed up the process by eliminating the need to investigate the parameters within the second group.

Partial contribution to variance

A number of studies in the literature suggest an approach of considering the partial contribution of each parameter to the variance in the outcome variable as a means of ranking the importance of the variables in a model.[68] The method involves calculating, for each parameter, the partial correlation coefficient (PCC) between that parameter and the outcome variable. The PCC measures the correlation between variables after adjusting for any indirect correlations with other variables. It is a parametric method and so assumes some linearity between the input parameters and the outcome. It excludes the influence of other parameters in the model by holding them constant. For example, if two variables, x_1 and x_2, are highly correlated with each other, and also both highly correlated with a third variable y, the PCC between x_1 and y, excluding the effect of x_2, would be

$$r_{x_1 y.x_2} = \frac{r_{x_1 y} - r_{x_2 y} r_{x_1 x_2}}{\sqrt{(1 - r^2_{x_1 x_2})}\sqrt{(1 - r^2_{x_2 y})}}$$

where $r_{x_1 y}$ = Pearson correlation coefficient of x_1 and y; $r_{x_2 y}$ = Pearson correlation coefficient of x_2 and y; and $r_{x_1 x_2}$ = Pearson correlation coefficient of x_1 and x_2.

By adjusting for the effect of x_1 when calculating the correlation between x_2 and y, the expected result would be a reduced correlation coefficient compared with the Pearson correlation coefficient. Once the PCC has been calculated for each parameter (excluding the effect of all other parameters in the model), the PCC values are squared to give the partial contribution to variance (PCV). This is done so that all parameters take positive values, and the magnitude of the value, rather than the sign of it, becomes important. The parameters are then ranked according to their associated PCV statistic, with parameters with the highest values being deemed the most important in the model.

Use of partial contribution to variance in EVI analysis

This method is potentially very useful in EVI analysis because of the likelihood of there being high correlations between the input variables. For example, there could be expected to be high correlations between various cost parameters and between utility parameters. This method would eliminate such correlations and therefore give an unbiased estimate of each parameter's contribution to the variance of the outcome variable (ENB). This would highlight not only the most valuable variables in the model, but also those which contribute very little to the variance of the outcome variable; this would provide input into metamodelling, and also allow further research to focus on relevant parameters.

Partial rank correlation coefficient (PRCC)

This method is a non-parametric equivalent of the PCV method described above, and uses ranks rather than actual data values in computing the correlation coefficient. Like the PCV method, the correlation coefficient between each input parameter and the outcome measure is adjusted for the effect of all other input parameters in the model. This method is appropriate when there are likely to be non-linear relationships between the inputs and output of the simulation model. In order to determine the order of importance of the input parameters $x_1 - x_j$ on their impact on the uncertainty in the outcome variable y, a number of steps are required:

1. Rank the values of each input parameter and the values of the outcome measure.
2. Compute the PRCC between each x parameter and the outcome measure, y, adjusting for the effects of the remaining x_{j-1} input parameters, as with the PCV method.
3. Take the absolute values of the resulting PRCCs.
4. Rank the parameters according to their absolute PRCCs (from largest to smallest).

The parameters with the highest absolute PRCCs are the most important in explaining the uncertainty in y.

Use of partial rank correlation coefficients in EVI analysis

This method would be useful in much the same way as the PCV method in assisting EVI analysis. Estimating the PRCC between each parameter and the ENB would again identify the most important parameters. This method could be expected to be more appropriate than using the unadjusted rank correlation coefficient, because of the likelihood of there being non-negligible correlations between input parameters.

Standardised regression coefficients

Regression analysis can be used as a method of ranking variables on their importance. By considering the magnitude of the regression coefficient of each parameter in the model, the parameters can be ranked according to this. However, the absolute coefficients computed cannot be used for this purpose because they are relative functions of the relative magnitude of each parameter. For this reason, the standardised regression coefficients must be used. The standardised coefficient, β_i, is given automatically for each parameter by many software packages including SAS® and SPSS®, but can also be calculated directly from the absolute regression coefficient as follows:

$$\beta_k = b_k \frac{\sigma_{xk}}{\sigma_y}$$

where b_k = the absolute coefficient of parameter k, σ_{xk} = the standard deviation of variable k and σ_y = the standard deviation of the outcome variable y.

The absolute value of each standardised coefficient is used so that the the coefficients of all parameters can be directly compared against each other. The parameters are then ranked according to their absolute standardised coefficient, with parameters having higher coefficients being the most important in the model, and vice versa.

Use of standardised regression coefficients in EVI analysis

To use this method for EVI analysis, one simply runs a regression with ENB as the dependent

variable and all input parameters as the independent variables. In SPSS, this can be done using the 'enter' method, which ensures that all independent variables are entered into the model initially. The absolute value of the standardised coefficient of each parameter is then computed, and the parameters are then ranked according to the size of this statistic, in descending order, with the parameters with the highest coefficients being the most important in the model.

Frequency domain methodology

Frequency domain methodology (FDM) differs from other importance analysis techniques because it requires only two runs of the simulation model to prioritise the parameters, irrespective of the number of input parameters. This is because the run length, as opposed to the number of runs, is increased to accommodate more input parameters.[71] The method computes a signal-to-noise ratio (SNR) for each parameter, derived from the results of the two simulation runs. This statistic is then compared against an *F*-statistic; the result of this comparison determines whether or not the parameter is significant in the model.

As mentioned above, two independent model runs are required:

- a signal run
- a noise run.

The signal run involves running the simulation model with the input parameters being varied cosinusoidally according to the following equation:[71]

$$X_j(t) = X_j(0) + a_j \cos(\omega_j t)$$

where t = the observation index (usually time), $X_j(0)$ = the mean value of parameter j, a_j = the oscillation amplitude and ω_j = the frequency in radians per unit of t.

Each input factor $X_j(t)$ is assigned a unique frequency ω_j, known as a driving frequency. These driving frequencies are part of a set called the Fourier frequencies, which are given by[71]

$$\omega_1 = \frac{2\pi p_j}{N}$$

where p_j = the parameter index number and N = the sample size.

A second simulation run (the noise run) is required to obtain an estimate of the variance of the error process, with the input parameters being held fixed so that the variation in the output is due solely to noise in the system.[71] For each run, and for each parameter, a statistic called the periodogram ordinate, $I(\omega_j)$, is calculated:

$$I(\omega_j) = \left(\frac{2}{N}\right)\left|\sum_{t=0}^{N-1} Y(t)\mathrm{e}^{-i\omega_j t}\right.$$

where $Y(t)$ = the outcome value for the model run.

In order to test the significance of the parameters, an SNR is computed for each parameter by taking the ratio of the periodograms generated from the signal and noise runs. The purpose of deriving this statistic is to remove an unknown nuisance parameter associated with the error process. The SNR is given by

$$\mathrm{SNR}(\varpi_j) = I^{\mathrm{s}}(\varpi_j)/I^{\mathrm{n}}(\varpi_j)$$

where I^{s} = the periodogram ordinate for the signal run and I^{n} = the periodogram ordinate for the noise run.

The distribution of the SNR has been shown to be well approximated by a non-central *F* distribution, $F_{2,2,\delta I}$. The test of the SNR has two hypotheses: the null hypothesis is that the effect of the parameter is negligible and unimportant; the alternative hypothesis is that the effect is not negligible. The SNR of each parameter is compared against the *F*-statistic. If the SNR for any given parameter is greater than the value of $F_{2,2}$, the null hypothesis is rejected in favour of the alternative, that is, the SNR for each parameter is compared against the *F*-statistic; parameters with an SNR statistic greater than $F_{2,2}$ are considered to be significant in the model. The reliability of this classification of parameters into two categories depends on the significance level used in the test. For example, the use of a 5% significance level would lead to more parameters being classed as important, whereas a 1% level would prove more stringent.

Use of frequency domain methodology in EVI analysis

To apply FDM to EVI analysis, a signal and a noise run would be carried out from the simulation model, from which an SNR could be calculated for each parameter. Each SNR would then be compared against the *F*-statistic mentioned above to determine whether each individual parameter was significant or not. Once these significant parameters have been identified, the size of their associated SNR could be used to rank their importance.

The issue of a suitable significant level to use in the *F*-test is important here, because the use of a more relaxed level such as 10% increases the chance of incorrectly classifying parameters as important when they are in fact not. The choice of the significance level is ultimately determined by the level of simplicity required in the metamodel (i.e. how many parameters can be handled by the metamodel). The best approach may be to begin by testing at a moderate significance level (e.g. 10%), and determining how many parameters are significant at this level. If it is considered that there are too many significant parameters, a more stringent significance level could be used to remove some of the less significant parameters.

Elasticity measures

Elasticity is a measure of the change in the value of the model outcome to a change in the value of an input parameter.[68] The elasticity, E_j, of the outcome to a given input parameter can be shown to be

$$E_j = \frac{\mathrm{d}y}{\mathrm{d}x_i}\frac{x_i}{y}$$

where x_i is a given input parameter, $\mathrm{d}x_i$ is the change in parameter x_i, y is the outcome measure and dy is the change in the output measure.

Point elasticity (i.e. at a single value) is an inappropriate measure because it reflects neither the level of uncertainty around a parameter nor the fact that elasticity will vary over the range of values of an input parameter.[68] These two types of uncertainty must therefore be taken into account simultaneously. Two alternative elasticity methods have been used as importance analysis measures:

1. actual elasticity coefficients (AECs)
2. absolute relative overall sensitivity (AROS).

AEC is the product of the point elasticity associated with an input parameter and its coefficient of variation; however, it does not take into account variability in elasticity over the range of input parameter values.[68] AROS estimates the responsiveness of the outcome to values for the input parameters by regression analysis[68] and incorporates the two types of uncertainty mentioned above. Each regression coefficient is used to make an estimate of the elasticity across a limited range of that parameter. AROS, however, relies on the assumption that the relationships between the input and output parameters are linear.

An approach which avoids this problem is the use of the elasticity coefficient (EC). This is estimated by repeating the Monte Carlo simulation while employing different fixed values for each parameter. The EC is then calculated for each parameter by the integration of elasticity values over all possible values of the parameter (x_p) weighted by its probability density function, $f(x_p)$:[68]

$$\mathrm{EC}(x_p) = \int_{-\varpi}^{\varpi} f(x_p).\frac{\mathrm{d}y}{\mathrm{d}x}.\frac{x_p}{y}\,\mathrm{d}x_p$$

Use of elasticity measures in EVI analysis

The application of elasticity methods to EVI analysis could be done in one of two ways. If the simulation model under consideration was known to be linear, AROS could be used as an approximation method for estimating the ENB of each treatment. This would involve carrying out linear regression to obtain estimates of the coefficients for each parameter, using ENB as the response variable. These coefficients could then be used to approximate the elasticity of the net benefit to each individual parameter.

If the simulation model was known to be non-linear, then the EC method could be adopted. Monte Carlo simulation would be used to run the model at a variety of values for each parameter. These different parameter values would then be used to derive a set of elasticity values (and a probability density function) for each parameter. The EC for each parameter could then be calculated via integration. The main issue in using this method would be in determining a suitable number of Monte Carlo simulation runs to perform, in order to provide elasticity values over all possible values of each parameter.

Conclusions

Importance analysis has been shown to be an important component of the metamodelling process, reducing and prioritising the number of parameters and thus making the metamodel simpler to build and interpret. Removing variables from a simulation model will have some effect on the accuracy of the metamodel, the magnitude of this effect being determined by the importance of the variables left out of the metamodel. One problem common to all importance analysis methods is determining how influential a parameter has to be to be considered important – this is ultimately subjective, and depends on the level of accuracy required from the metamodel in approximating the simulation model.

Correlation- and variance-based methods tend to be the most widely used methods of importance analysis owing their relative simplicity, but some

other relatively straightforward techniques have been presented here which offer alternatives to these techniques. As with metamodelling techniques, no single method is appropriate in all situations. SB is the most useful when only a small number of parameters are thought to be important, and is flexible in exploring the effects of different parameters,[72] but is not so effective with a larger number of important parameters. Because of the advantages and disadvantages of the different methods, it is recommended to use more than one method to prioritise variables; a comparison of the parameters selected by these methods could then be made to see whether the results were consistent. A similarity between the parameters selected from these different methods would generally give greater confidence in the subsequent metamodel.

This chapter has identified and critically explored the potential value of a number of alternative metamodelling techniques that may be used to reduce computation time in undertaking EVI analysis. However, for several of these methods, the body of literature is weak and incomplete, hence the merit of these techniques and accessibility for a general user of health economic models can at best be only suggested. The potential role of these methods is therefore further explored in the following chapter through the direct application to the case study ScHARR MS model.

Chapter 4

Applied methodology: EVI analysis for computationally expensive health economic models

Introduction

This chapter describes the direct application of the methodological framework put forward in Chapter 3 in order to perform EVPI analysis for the case study model of IFN-β and glatiramer acetate in the management of MS. This chapter includes a direct comparison of the global and partial EVPI results from three models:

- the public domain ScHARR MS cost-effectiveness model
- a multiple linear regression metamodel to approximate the ScHARR model
- a GP metamodel to approximate the ScHARR model.

The EVPI analysis presented within this chapter assumes a maximum acceptable incremental cost-effectiveness ratio of £30,000. This chapter reports only on 'per patient EVPIs'; population EVPIs are presented in Chapter 5.

Can EVI be calculated numerically?

The method outlined in the section 'Can EVPI be calculated numerically' (p. 17) was applied to the cost H parameter within the MS model. The method is demonstrated using this parameter as the importance analysis suggested that considerable uncertainty was explained by the cost H parameter [see the section 'Factor screening to identify important variables in the model' (p. 46)]. For estimating these sample sizes, a potentially key variable should be selected either from prior knowledge of the model structure or through undertaking a preliminary linear regression on the ENBs of treatment strategies using the Monte Carlo sample for estimating the overall EVPI. The sample size s was increased sequentially to check for convergence, as shown in *Table 6*.

TABLE 6 *Mean bias with varying sample size*

n	*s* = 30	*s* = 40	*s* = 50	*s* = 60
500	456	423	431	613
1,000	214	218	183	318
5,000	45	25	41	75
10,000	20	6	23	41

Although the estimates of the mean bias have not converged, there is a clear indication here that an inner sample size of 1000 runs is likely to produce a substantial bias, and obtaining a two-level Monte Carlo estimate of the partial EVPI for cost H is likely to require severe computational effort.

The initial set of 620 runs can also be used to estimate an appropriate outer sample size. Denoting the outer sample size by r, an approximate 95% confidence interval (CI) for the partial EVPI is given by

$$\text{EVPI}(\theta_i) \pm 1.96\,[\text{Var}\{h(\theta_i)\}/r]^{0.5}$$

with $h(\theta_i)$ the Monte Carlo estimator of $E\theta_{-i}\{\max NB(d, \theta)|\theta_i\}$. We write

$$h(\theta_i) = E\theta_{-i}\{\max NB(d, \theta)|\theta_i\} + \varepsilon$$

that is, $h(\theta_i)$ is the true value of the expectation plus a random error resulting from Monte Carlo sampling. The error term ε will not have expectation zero as the Monte Carlo estimator is biased. We will then make the simplifying assumption

$$\text{Var}[h(\theta_i)] = \text{Var}\,[E\theta_{-i}\{\max NB(d, \theta)|\theta_i\}] + \text{Var}(\varepsilon)$$

The variance of ε may depend on θ_i, but for the purpose of this sample size calculation we shall assume the variance is constant. The first term is the variance taken with respect to θ_i, and can be estimated using numerical integration; at 21 values of θ_i we have an estimate of $E\theta_{-i}\{\max NB(d, \theta)|\theta_i\}$ based on 30 samples of θ_{-i}. We can also inspect Var(ε) for different values of θ_i by calculating the variance of the Z values generated when investigating the bias for inner sample sizes

TABLE 7 *Confidence interval width for partial EVPIs at various sample sizes*

Sample size *r*	100	1000	10,000	100,000
CI width	11,729	3709	1173	371

(we will have one sample of Z values for each of the 21 values of θ_i). Note that Var(ε) will decrease as the inner sample size increases.

Returning to the example, suppose we have decided on an inner sample size of 1000 for the inner sample size. We then obtain the following estimates of the two variance components:

$$\text{Var}\left[E\theta_{-i}\left\{\max NB(d, \theta)|\theta_i\right\}\right] = 8.9528 \times 10^8$$

$$\text{Var}(\varepsilon) \in (0.3943 \times 10^4, 1.4417 \times 10^4)$$

For Var(ε) we have reported the minimum and maximum variance as θ_i varies. The dependency of the variance on θ_i does not matter here, as the overall variance is dominated by the first term. We can now estimate the width of a 95% CI for the partial EVPI for various outer sample sizes, as shown in *Table 7*.

Table 7 shows that the CI are fairly wide, indicating that a large outer sample size will also be required, certainly in excess of 1000 samples. In conclusion, on the basis of 620 model evaluations, it has been possible to establish that a likely minimum sample size for an accurate two-level Monte Carlo estimate will be of the order of 1,000,000 model runs. The calculations shown in *Table 1* suggest that 1,000,000 model runs for a single parameter would require approximately 81 days, which is clearly infeasible.

Is the model linear?

We performed multiple linear regression and correlation analysis using SPSS in order to test whether a linear relationship exists between the randomly sampled input model parameters and the NBs calculated using the ScHARR model, and to investigate the extent to which the variation in the NBs is explained by the variation in the parameter inputs. Multivariate Monte Carlo sampling was performed over 10,000 random iterations using a one-level sampling algorithm, allowing all parameters to vary according to their prior uncertainty. The absolute NB for all seven treatment strategies were calculated for each iteration. The sampled values for each model variable were imported into SPSS together with their associated expected net benefit across all 10,000 iterations.

The independent variables for the analysis were the randomly sampled parameter values. The dependent variable for each regression model was NB. Multiple linear regression analysis was performed for each treatment strategy to produce seven separate linear models. *Table 8* demonstrates that for each of the seven linear approximations, the adjusted r^2 is high (0.93 for all models), which suggests that there is a strong linear relationship between the behaviour of the distributions of parameter inputs and the resulting NBs. Although the model is not perfectly linear, the regression analysis indicates that ~93% of the variation in the distribution of NBs is explained by the parameter inputs for the model. First-order interactions were not included in this model as the software was unable to support the additional independent variables for the analysis (8384). The inclusion of first-order interactions within the regression analysis would require programming in a fast language, which would ultimately make the metamodelling process unnecessarily complex. However, even without the inclusion of these interaction terms, the adjusted r^2 statistic suggests a reasonable model fit.

Table 9 shows that the all seven models are significant, as demonstrated by the significance p values.

TABLE 8 *Results of the multiple linear regression analysis*

Treatment strategy	r^2	Adjusted r^2	Standard error of the estimate
T0 (Conventional management)	0.93	0.93	27717
T1 (Avonex)	0.93	0.93	26011
T2 (Rebif 22 μg)	0.93	0.93	26585
T3 (Rebif 44 μg)	0.93	0.93	26292
T4 (Betaferon RRMS only)	0.93	0.93	26673
T5 (Copaxone)	0.93	0.93	27692
T6 (Betaferon RRMS and SPMS)	0.93	0.93	26629

TABLE 9 *Overall significance of the linear models*

Treatment strategy	Sum of squares	Degrees of freedom	Mean square	*F*	Significance *p*
T0	1.06×10^{14}	103	1.03×10^{12}	1343	0.000
T1	8.92×10^{13}	105	8.49×10^{11}	1255	0.000
T2	9.10×10^{13}	105	8.66×10^{11}	1226	0.000
T3	8.75×10^{13}	105	8.33×10^{11}	1205	0.000
T4	9.18×10^{13}	105	8.74×10^{11}	1228	0.000
T5	9.87×10^{13}	105	9.40×10^{11}	1226	0.000
T6	9.10×10^{13}	107	8.51×10^{11}	1200	0.000

TABLE 10 *Global EVPI results for the ScHARR MS model for each individual treatment strategy*

	EVPI T1 vs T0	EVPI T2 vs T0	EVPI T3 vs T0	EVPI T4 vs T0	EVPI T5 vs T0	EVPI T6 vs T0	Decision EVPI
Per patient EVPI (£)	4271	3035	2827	2776	2444	3514	8855

Undertaking EVPI analysis using the original case study ScHARR model

Although it was not feasible to perform a full EVPI analysis for individual parameters within the ScHARR MS model using the two-level sampling algorithm, the strong linear relationship between the sampled parameter values and the NBs for each treatment strategy suggests that the one-level algorithm may provide a reasonable approximation of the true EVPI associated with each parameter within the decision model.

Global EVPI analysis using the original ScHARR MS model

The global EVPI was calculated analytically using the original ScHARR MS model, thus representing the most realistic estimate of the value of information across all parameters to the decision problem. Multivariate Monte Carlo sampling was undertaken using the ScHARR MS model in order to estimate the uncertainty across the entire design space, allowing all parameters within the model to vary simultaneously. A total of 10,000 iterations were performed to ensure stability in the EVPI estimates. This algorithm, which is described in *Box 1* (p. 5), requires only a one-level sampling method as the calculation involves only a single expectation. For each iteration, the absolute costs and QALYs for each treatment strategy were recorded. [Note: these data are the same as those used to test the linearity of the model, as described in the section 'Is the model linear?' (p. 40)]. The absolute NBs of each treatment strategy were then calculated, together with the maximum NB for each iteration (the NB of the optimal strategy). The overall EVPI was calculated as the average of the maximum NBs minus the maximum of the average NBs [see *Box 1* (p. 5)].

The results of the global EVPI analysis are presented in *Table 10* for each individual treatment strategy (i.e. DMT versus conventional management); the lower estimate for the overall per patient EVPI is £4271. The overall EVPI across all treatment strategies is £8855; this represents the upper estimate of the overall EVPI.

Partial EVPI analysis using the original ScHARR MS model

The global EVPI calculated above suggests that there is considerable value in collecting further information relevant to this decision problem. The global EVPI, however, does not provide an indication of which aspects of the decision further information is expected to yield the greatest value. Partial EVPI analysis for groups of parameters and individual parameters is thus used to identify which parameters are most important to the decision problem.

Partial EVPI analysis was performed using the ScHARR MS model to identify which parameters are likely to be most influential in reducing current decision uncertainty. Owing to the vast computational expense involved in performing a full two-level analysis using the ScHARR MS model, it was only feasible to apply a one-level sampling design for groups of parameters.

The algorithm for performing one-level EVPI analysis is described in full in *Box 3* (p. 7). Monte Carlo sampling was undertaken using a one-level sampling technique via the ScHARR model, allowing those parameters of interest to vary whilst

all other parameters not of interest were held at their mean values.

Table 11 shows the ENBs for each treatment strategy and hence the resulting partial EVPI results for groups of parameters within the model (a description of the groups of parameters in the ScHARR model is shown in Appendix 2). It is clear that EDSS costs, utilities, relapse parameters and the relative risks of treatment effect are expected to yield the greatest value of information.

Table 11 demonstrates that for several groups of parameters, the collection of further information is expected to yield no value, hence undertaking partial EVPI analysis for the individual parameters within these groups would similarly be expected to result in zero value. Undertaking partial EVPI analysis for groups of parameters prior to undertaking partial EVPI analysis for all individual parameters may therefore help to identify those parameters for which further research is expected to be valuable and those for which further research is not expected to be valuable, thus considerably reducing computation time required for the analysis.

Table 12 shows the partial EVPI results for those individual parameters for which further data collection is expected to yield some value, as identified from the results of the partial EVPI for groups of parameters. It is evident that most value is likely to be obtained through further research on utility parameters *A* and *C*, cost parameter *H*, the cost of relapse and the relative risks of disease progression for those patients with RRMS. The results also suggest that there is some value in obtaining further information concerning the rate at which patients drop off therapy. These results are clearly consistent with the partial EVPI estimates for groups of parameters shown in *Table 11*.

Practical critique of the one-level EVPI algorithm

Does it save time?

The use of the one-level sampling algorithm in place of the two-level approach has potential to dramatically reduce the computation time required to perform a comprehensive value of information analysis. The one-level algorithm used to calculate the partial EVPI for each individual or group of parameters is almost identical to the global EVPI algorithm; the only methodological difference in undertaking partial EVPI analysis for individual and groups of parameters using a one-level sampling algorithm concerns which parameters are allowed to vary and which parameters are held at their mean values. The actual calculation of partial EVPI is otherwise equivalent to the calculation of global EVPI. Hence the computation time required to calculate the partial EVPI for all individual parameters is simply the time required to run the required number of random simulations multiplied by the number of parameters within the model. The analysis presented here assumes a MAICER of £30,000; performing EVPI analysis

TABLE 11 *Partial EVPI for groups of parameters calculated using the ScHARR MS model one-level algorithm (all data in £)*

Parameter group	ENB T1	ENB T2	ENB T3	ENB T4	ENB T5	ENB T6	ENB T0	Maximum net benefit	EVPI for parameter group
1. EDSS costs	184,453	181,637	178,748	180,964	175,980	180,645	193,270	195,301	2,030
2. EDSS utilities	187,164	184,029	181,282	183,288	177,849	183,249	194,790	195,783	993
3. Relapse – duration, cost and disutility	190,646	188,123	185,158	187,491	182,733	187,236	200,027	200,716	688
4. EDSS duration – beta sojourn times	191,242	188,673	185,691	188,044	183,314	187,765	200,721	200,721	0
5. Relative risks of treatment effect	191,630	188,805	185,867	188,160	183,249	187,999	200,812	203,763	2,951
6. Side-effects	191,404	188,802	185,823	188,166	183,388	187,893	200,812	200,812	0
7. Dropouts	191,053	188,538	185,496	187,916	183,214	187,520	200,812	200,967	155
8. Relapse count	191,441	188,829	185,850	188,192	183,408	187,920	200,841	200,841	0
9. Mean sojourn times in EDSS states	191,169	188,588	185,602	187,955	183,192	187,671	200,621	200,621	0
10. DSS to EDSS ratios	191,022	188,359	185,405	187,711	182,855	187,498	200,224	200,224	0

TABLE 12 *Partial EVPI estimates using the ScHARR MS model 1-level algorithm (all data in £)*

Parameter	ENB T1	ENB T2	ENB T3	ENB T4	ENB T5	ENB T6	ENB T0	Maximum ENB	EVPI for individual parameter
Utility parameter *A*	191,520	188,923	185,941	188,288	183,512	188,015	200,948	201,264	316
Utility parameter α	197,572	195,047	192,047	194,416	189,616	194,287	206,993	206,993	0
Utility parameter *C*	179,959	176,664	173,972	175,904	170,399	175,745	187,346	187,764	418
Disutility of relapse	191,436	188,830	185,850	188,194	183,414	187,919	200,855	200,855	0
Cost parameter *L*	191,368	188,767	185,787	188,131	183,352	187,857	200,787	200,787	0
Cost parameter *H*	191,390	188,788	185,808	188,152	183,372	187,879	200,806	201,687	881
Cost of relapse	192,069	189,396	186,403	188,757	183,960	188,467	201,514	202,118	603
Cost parameter *R*	184,660	181,860	178,960	181,191	176,233	180,846	193,564	193,564	0
T1 RRMS relative risk of progression	191,588	188,792	185,813	188,156	183,377	187,952	200,812	201,307	495
T2 RRMS relative risk of progression	191,394	188,758	185,813	188,156	183,377	187,952	200,812	201,334	522
T3 RRMS relative risk of progression	191,394	188,792	185,923	188,156	183,377	187,952	200,812	200,956	144
T4 RRMS relative risk of progression	191,394	188,792	185,813	188,173	183,377	187,952	200,812	201,301	489
T5 RRMS relative risk of progression	191,394	188,792	185,813	188,156	183,512	187,952	200,812	201,787	975
T6 RRMS relative risk of progression	191,394	188,792	185,813	188,156	183,377	188,072	200,812	201,271	459
T6 SPMS relative risk of progression	191,394	188,792	185,813	188,156	183,377	187,939	200,812	200,812	0
T1 RRMS relative risk of relapse	191,407	188,792	185,813	188,156	183,377	187,952	200,812	200,812	0
T2 RRMS relative risk of relapse	191,394	188,771	185,813	188,156	183,377	187,952	200,812	200,812	0
T3 RRMS relative risk of relapse	191,394	188,792	185,814	188,156	183,377	187,952	200,812	200,812	0
T4 RRMS relative risk of relapse	191,394	188,792	185,813	188,159	183,377	187,952	200,812	200,812	0
T5 RRMS relative risk of relapse	191,394	188,792	185,813	188,156	183,385	187,952	200,812	200,812	0
T6 RRMS relative risk of relapse	191,394	188,792	185,813	188,156	183,377	187,872	200,812	200,812	0
Relapse duration	191,388	188,787	185,807	188,151	183,372	187,878	200,806	200,806	0
Proportion of patients experiencing side effects	191,392	188,790	185,811	188,154	183,375	187,881	200,812	200,812	0
Side effects utility adjustment	191,383	188,781	185,801	188,145	183,366	187,871	200,812	200,812	0
Year 1, 2 dropouts	191,295	188,715	185,719	188,083	183,325	187,785	200,812	200,941	129
Subsequent dropouts	191,191	188,644	185,627	188,017	183,285	187,652	200,812	200,819	7

over a range of different thresholds does not involve any additional computational time, as the ENBs can be calculated outside of the model itself.

For the case study model, the calculation of the global EVPI for the case study model took ~19 hours, the calculation of partial EVPI for the 10 parameters groups took ~194 hours, whereas a full partial EVPI analysis for all 128 individual parameters using the one-level algorithm would require ~2489 hours of simulation time. However, for the case study model the EVPI analysis for groups of parameters (see *Table 11*) showed that the collection of further information for several parameter groups is expected to yield no further value, thus reducing the number of individual parameters requiring analytical investigation. As a result, partial EVPI analysis was performed for 16 parameters contained within these groups; this required ~311 hours of computation time. Hence undertaking partial EVPI analysis for groups of parameters can be instrumental in the identification of those parameters for which further data collection is likely to yield some value and those for which further data collection would not.

The potential reduction in computation of the one-level EVPI approach should not be understated; assuming that 10,000 samples were required for both the inner and outer level sampling within the case study model, the equivalent full EVPI analysis using the two-level algorithm would require over 2800 years over all individual parameters. A lesser payoff in terms of computation time saved would be achieved through using the one-level algorithm for other health economic models that are less computationally expensive than the case study model.

Are there parametric restrictions for this method?
As partial EVPI analysis for parameters is undertaken on a parameter-by-parameter basis, the only parametric restriction concerns the overall time available for the project. Owing to time constraints, it was not possible to perform partial EVPI analysis for all 128 individual parameters within the case study model. However, it is a crucial point that the results of the partial EVPI for the groups of parameters suggested that many of these simulations would in fact be unnecessary. It should be noted that for models of extreme computational expense, such as the ScHARR MS model, the Windows 98 operating system may have insufficient memory capacity to support full Monte Carlo sampling algorithms over a large number of random iterations. The Windows XP operating system, however, caches hard memory using an alternative process, thus resolving this problem.

Closeness of results
The global EVPI calculated using the original model should be taken to be the comparator for the different methodologies. With regard to the one-level partial EVPI evaluations for parameters, the payoff achieved in terms of reduced computational expense may be offset by reduced accuracy of results. The degree to which the partial EVPI estimates are biased will be dependent on the strength of the linear relationship between the sampled parameter inputs and NBs. The regression analysis presented in the section 'Is the model linear?' (p. 40) suggests that a reasonably strong linear relationship exists between the parameter inputs and the NBs within the MS model; however, as the EVPI calculation uses the extreme values of the sampled data, the sampling bias is likely to have distorted these EVPI estimates.

Ease of implementation
The calculation of the global EVPI requires a one-level sampling algorithm via multivariate Monte Carlo simulation, which may be easily implemented using VBA alongside EXCEL or using other dedicated decision analysis software. The calculation of both global EVPI and partial EVPI for parameters from the results of the Monte Carlo sampling is very straightforward, as illustrated in *Table 2*.

Is the method algorithmic or heuristic?
The one-level method is predominantly algorithmic. The development of methods for handling uncertainty in health economic models means that probabilistic Monte Carlo simulation has become commonplace within the field of health economic modelling and requires little further specialist expertise beyond some basic knowledge of VBA programming or alternative decision analysis software.

Undertaking EVPI analysis using a multiple linear regression metamodel

The use of the one-level sampling algorithm using the original MS model resulted in considerable computational expense, thus making even one-level EVPI analysis for all model parameters infeasible. Due to the reasonable

linear relationship between the model inputs and the NBs for each treatment strategy, we investigated the feasibility of replacing the original cost-effectiveness model with a statistical approximation, whereby each treatment strategy is described by a regression equation [see the section 'Simple linear regression' (p. 21)].

The results of the regression analysis [see the section 'Is the model linear?' (p. 40)] were recorded together with the non-standardised β coefficients for each parameter within each of the seven regression models. These coefficients were then applied to each of the Monte Carlo iterations calculated previously. The primary advantage of this form of metamodelling is that full EVPI analysis (global EVPI, partial EVPI for groups of parameters and partial EVPI for all individual parameters) using the one-level sampling algorithm could be performed in < 1 hour of computation time. In addition, the development of such a regression model is straightforward via spreadsheet packages. In EXCEL, each model can be generated using a =SUMPRODUCT(array1, array2) function, where array 1 is the vector of non-standardised β coefficients and array 2 is the sampled parameter input vector for each random iteration.

Table 13 reports the global EVPI results obtained using the one-level algorithm via the linear regression metamodel. It should be noted that the approximated global EVPI is £1500 lower than the global EVPI calculated using the actual MS cost-effectiveness model. This lower estimate is likely to be a result of the absence of parameter interactions within the regression model and the imperfect linear relationship between sampled model inputs and net benefits.

Table 14 presents the partial EVPI results for the 10 groups of parameters using the one-level algorithm within the linear regression metamodel. These partial EVPI results are broadly consistent with the partial EVPIs for the groups of parameters calculated using the ScHARR MS model via the one-level algorithm. It is noteworthy that the magnitude of estimated EVPI for each group (except the relapse parameters) is broadly similar across the two methods. The sum of these

TABLE 13 *Global EVPI calculated using the multiple linear regression metamodel*

	EVPI T1 vs T0	EVPI T2 vs T0	EVPI T3 vs T0	EVPI T4 vs T0	EVPI T5 vs T0	EVPI T6 vs T0	Decision VPI
Per patient EVPI (£)	3740	2385	2186	2217	1919	3074	7263

TABLE 14 *Partial EVPI for groups of parameters calculated using the one-level algorithm within the multiple linear regression metamodel (all data in £)*

Parameter group	ENB T1	ENB T2	ENB T3	ENB T4	ENB T5	ENB T6	ENB T0	Maximum ENB	EVPI for parameter group
1. EDSS costs	184,295	177,378	173,956	171,336	173,210	167,392	173,176	186,296	2,001
2. EDSS utilities	188,461	181,047	177,865	175,178	177,155	171,504	177,163	189,208	747
3. Relapse – duration, cost and disutility	185,307	178,197	174,889	172,250	174,160	168,414	174,127	185,307	0
4. EDSS duration – beta sojourn times	185,314	178,215	174,895	172,249	174,170	168,441	174,141	185,314	0
5. Relative risks of treatment effect	185,287	178,376	174,709	171,949	174,307	168,569	174,244	189,748	4,461
6. Side-effects	185,288	178,179	174,872	172,234	174,143	168,399	174,112	185,288	0
7. Dropouts	185,286	178,235	174,918	172,286	174,187	168,437	174,151	185,599	312
8. Relapse count	185,224	178,125	174,799	172,178	174,066	168,328	174,036	185,224	0
9. Mean sojourn times in EDSS states	185,415	178,298	174,992	172,351	174,265	168,523	174,219	185,415	0
10. DSS to EDSS ratios	185,257	178,150	174,830	172,200	174,114	168,361	174,070	185,257	0

partial EVPIs is around £7500, which is comparable with the overall estimate of global EVPI calculated using the same regression model; the consistency of results is again due to the absence of interactions between the model parameters; hence these results are likely to underestimate the true EVPI for each parameter group.

Table 15 shows the results of the partial EVPI analysis for all parameters estimated via the one-level sampling algorithm, as calculated using the linear regression model. Again the broad consistency between these EVPI estimates and those calculated via the ScHARR MS model (see *Table 12*) is noteworthy.

Table 15 shows that the key areas where further research is expected to yield the most value are the relationship between the EDSS, costs of care and health outcomes, the rate at which patients drop off therapy and the impact of therapy on the progression of the disease.

Practical critique of the linear regression metamodel

Does it save time?

The key advantage of using regression metamodelling to undertake value of information analysis is the modest computation time required to perform a comprehensive evaluation across all individual parameters within a decision model. For the case study ScHARR MS model, the analysis took a total of ~1 hour to approximate the global EVPI, the partial EVPI analysis for all 10 groups of parameters and also a complete partial EVPI approximation for all 128 individual model parameters. The computation time would be reduced further for decision models containing a smaller number of parameters.

Are there parametric restrictions for this method?

There are no parametric restrictions for this method. However, for health economic models which contain a large number of parameters, a greater number of observations are required. For the case study model, 10,000 samples were generated for each of the 128 parameters to the model, which was assumed to be sufficient to cover the entire range of uncertainty. This, however, almost exhausted the memory capacity of the computer on which the analysis was performed, which suggests that larger models may potentially encounter technical problems in generating the required number of samples. If the original cost-effectiveness model contains a large number of parameters, similar problems may arise if first-order interactions are specified in the regression metamodels.

Closeness of results

As with the application of the one-level EVPI algorithm using the original ScHARR model, the reduced computation time may potentially carry a cost in terms of accuracy of results. The extent to which the EVPI results are distorted is dependent on the degree of linearity between the model inputs and output. If one compares the global EVPI results calculated using the original ScHARR model with those approximated by the regression model, it is clear that the regression model underestimates the true global EVPI. It is reasonable to assume that the model also underestimates the partial EVPI results for individual parameters. This effect is likely to be due to the imperfect linear relationship between sampled inputs and associated NBs, together with the assumption of independence of model parameters within the regression model, hence interactions between model variables observed within the original ScHARR model are lost when the EVPI analysis is undertaken using the regression metamodel.

Ease of implementation

As suggested in Chapter 3, an important advantage of regression-based metamodelling is that the technique is very straightforward to implement. Within the case study, the original ScHARR model was used to generate 10,000 random samples of input parameters. These were imported into SPSS in order to calculate regression coefficients for each parameter. The Monte Carlo simulation routine was then replicated for the metamodel using EXCEL and VBA and combined with the regression coefficients. This process was undertaken for all seven treatment strategies. This process, including the programming of VBA subroutines to estimate partial EVPI for parameters, took ~2 days.

Is the method algorithmic or heuristic?

Multiple linear regression analysis is again a technique commonly applied within the field of health economics and modelling, although as with the one-level EVPI method, some specialist programming expertise may be required to implement the VBA subroutines.

Factor screening to identify important variables in the model

As identified in Chapter 3, the GP methodology is constrained by the number of parameters that can

TABLE 15 *Partial EVPI for individual parameters calculated using the one-level algorithm within the linear regression metamodel (all data in £)*

Parameter	ENB T1	ENB T2	ENB T3	ENB T4	ENB T5	ENB T6	ENB T0	Maximum ENB	EVPI for individual parameter
Probability female	185,294	178,182	174,879	172,238	174,152	168,406	174,117	185,294	0
Cost parameter *L*	185,253	178,143	174,836	172,198	174,108	168,364	174,076	185,253	0
Cost parameter *H*	185,554	178,369	175,085	172,436	174,360	168,646	174,307	186,594	1,040
Cost parameter *R*	183,592	176,767	173,370	170,774	172,625	166,787	172,626	183,592	0
Cost of relapse	185,757	178,632	175,279	172,627	174,545	168,788	174,500	186,443	685
Utility parameter *A*	189,276	181,861	178,683	175,996	177,975	172,316	177,979	189,662	386
Utility parameter α	184,787	177,607	174,319	171,670	173,593	167,879	173,550	184,787	0
Utility parameter β	184,973	177,934	174,605	171,978	173,873	168,104	173,855	185,017	44
Disutility of relapse	185,334	178,224	174,913	172,273	174,182	168,437	174,150	185,334	0
Relapse duration	185,205	178,092	174,803	172,159	174,069	168,335	174,043	185,205	0
RR0	185,299	178,189	174,884	172,245	174,154	168,411	174,123	185,299	0
RR1	185,283	178,173	174,867	172,227	174,136	168,393	174,106	185,283	0
RR1.5	185,293	178,188	174,877	172,241	174,148	168,405	174,116	185,293	0
RR2	185,292	178,181	174,871	172,236	174,148	168,399	174,115	185,292	0
RR2.5	185,298	178,186	174,882	172,243	174,152	168,409	174,123	185,298	0
RR3	185,292	178,181	174,875	172,237	174,146	168,401	174,115	185,292	0
RR3.5	185,283	178,176	174,863	172,225	174,139	168,399	174,102	185,283	0
RR4	185,291	178,179	174,870	172,232	174,144	168,400	174,114	185,291	0
RR4.5	185,242	178,143	174,838	172,191	174,103	168,363	174,077	185,242	0
RR5	185,287	178,178	174,871	172,233	174,143	168,398	174,111	185,287	0
RR5.5	185,285	178,176	174,868	172,230	174,140	168,396	174,109	185,285	0
RR6	185,294	178,185	174,880	172,240	174,150	168,405	174,117	185,294	0
RR6.5	185,292	178,183	174,875	172,237	174,146	168,404	174,113	185,292	0
RR7	185,293	178,180	174,875	172,237	174,149	168,403	174,115	185,293	0
RR7.5	185,282	178,174	174,866	172,229	174,138	168,393	174,106	185,282	0
RR8	185,286	178,175	174,870	172,231	174,141	168,397	174,109	185,286	0
RR8.5	185,289	178,180	174,873	172,234	174,144	168,398	174,112	185,289	0
RR9	185,302	178,188	174,884	172,243	174,151	168,415	174,124	185,302	0
RR9.5	185,305	178,199	174,888	172,248	174,164	168,409	174,129	185,305	0
SE prop. of people	185,288	178,177	174,871	172,232	174,142	168,397	174,110	185,288	0
SE utility adjustment	185,287	178,179	174,873	172,234	174,144	168,399	174,113	185,287	0
Year 1, 2 dropouts	185,294	178,354	175,008	172,400	174,277	168,498	174,281	185,544	250
Subsequent dropouts	185,280	178,058	174,782	172,119	174,053	168,336	173,981	185,317	36
Relapse count RRMS0	185,282	178,171	174,866	172,227	174,138	168,393	174,106	185,282	0
Relapse count RRMS 1	185,315	178,214	174,899	172,262	174,172	168,422	174,139	185,315	0
Relapse count RRMS 1.5	185,292	178,183	174,876	172,237	174,147	168,402	174,115	185,292	0
Relapse count RRMS2	185,175	178,072	174,767	172,136	174,045	168,301	174,010	185,175	0
Relapse count RRMS2.5	185,336	178,221	174,910	172,270	174,178	168,438	174,150	185,336	0

continued

TABLE 15 *Partial EVPI for individual parameters calculated using the one-level algorithm within the linear regression metamodel (all data in £) (cont'd)*

Parameter	ENB T1	ENB T2	ENB T3	ENB T4	ENB T5	ENB T6	ENB T0	Maximum ENB	EVPI for individual parameter
Relapse count RRMS3	185,310	178,196	174,889	172,248	174,161	168,420	174,129	185,310	0
Relapse count RRMS3.5	185,294	178,183	174,875	172,237	174,148	168,402	174,116	185,294	0
Relapse count RRMS4	185,288	178,179	174,871	172,233	174,143	168,398	174,111	185,288	0
Relapse count RRMS4.5	185,252	178,149	174,845	172,200	174,110	168,358	174,079	185,252	0
Relapse count RRMS5	185,260	178,158	174,846	172,212	174,120	168,370	174,086	185,260	0
Relapse count RRMS5.5	185,289	178,173	174,865	172,229	174,142	168,397	174,107	185,289	0
Relapse count RRMS6	185,284	178,175	174,868	172,229	174,139	168,394	174,108	185,284	0
Relapse count RRMS6.5	185,291	178,182	174,875	172,238	174,145	168,402	174,114	185,291	0
Relapse count RRMS7	185,276	178,169	174,862	172,222	174,131	168,389	174,100	185,276	0
Relapse count RRMS7.5	185,312	178,198	174,895	172,254	174,165	168,425	174,134	185,312	0
Relapse count RRMS8	185,292	178,180	174,875	172,237	174,145	168,402	174,115	185,292	0
Relapse count RRMS8.5	185,285	178,175	174,868	172,231	174,141	168,397	174,106	185,285	0
Relapse count RRMS9	185,281	178,173	174,864	172,229	174,134	168,389	174,104	185,281	0
Relapse count RRMS9.5	185,281	178,173	174,867	172,228	174,137	168,391	174,105	185,281	0
Relapse count SPMS2	185,278	178,171	174,863	172,225	174,134	168,388	174,103	185,278	0
Relapse count SPMS2.5	185,285	178,176	174,869	172,231	174,141	168,396	174,109	185,285	0
Relapse count SPMS3	185,298	178,185	174,880	172,242	174,154	168,410	174,120	185,298	0
Relapse count SPMS3.5	185,298	178,187	174,881	172,243	174,151	168,406	174,121	185,298	0
Relapse count SPMS4	185,302	178,190	174,885	172,248	174,155	168,413	174,125	185,302	0
Relapse count SPMS4.5	185,279	178,169	174,863	172,225	174,135	168,388	174,102	185,279	0
Relapse count SPMS5	185,288	178,179	174,872	172,233	174,143	168,398	174,111	185,288	0
Relapse count SPMS5.5	185,288	178,178	174,872	172,234	174,144	168,398	174,112	185,288	0
Relapse count SPMS6	185,287	178,177	174,871	172,233	174,142	168,397	174,111	185,287	0
Relapse count SPMS6.5	185,276	178,167	174,862	172,222	174,133	168,386	174,102	185,276	0
Relapse count SPMS7	185,293	178,181	174,877	172,238	174,146	168,403	174,114	185,293	0
Relapse count SPMS7.5	185,295	178,183	174,878	172,239	174,149	168,404	174,118	185,295	0
Relapse count SPMS8	185,282	178,174	174,866	172,228	174,138	168,392	174,106	185,282	0
Relapse count SPMS8.5	185,283	178,173	174,867	172,228	174,137	168,393	174,106	185,283	0
Relapse count SPMS9	185,279	178,170	174,860	172,225	174,132	168,389	174,097	185,279	0
Relapse count SPMS9.5	185,289	178,179	174,872	172,235	174,143	168,398	174,112	185,289	0
RR0	185,331	178,224	174,918	172,279	174,188	168,442	174,156	185,331	0
RR1	185,236	178,151	174,837	172,201	174,107	168,355	174,077	185,236	0
RR2	185,323	178,191	174,893	172,249	174,163	168,423	174,132	185,323	0
RR3	185,266	178,166	174,858	172,220	174,127	168,382	174,097	185,266	0
RR4	185,292	178,179	174,874	172,235	174,145	168,402	174,113	185,292	0
RR5	185,282	178,175	174,868	172,229	174,139	168,393	174,107	185,282	0
RR6	185,274	178,172	174,862	172,224	174,135	168,388	174,102	185,274	0
RR7	185,287	178,177	174,870	172,232	174,142	168,397	174,110	185,287	0

continued

TABLE 15 *Partial EVPI for individual parameters calculated using the one-level algorithm within the linear regression metamodel (all data in £) (cont'd)*

Parameter	ENB T1	ENB T2	ENB T3	ENB T4	ENB T5	ENB T6	ENB T0	Maximum ENB	EVPI for individual parameter
RR8	185,308	178,196	174,890	172,250	174,161	168,416	174,131	185,308	0
RR9	185,318	178,200	174,897	172,257	174,169	168,429	174,139	185,318	0
SP2	185,279	178,170	174,863	172,227	174,134	168,389	174,103	185,279	0
SP3	185,329	178,218	174,910	172,271	174,180	168,435	174,143	185,329	0
SP4	185,295	178,182	174,877	172,238	174,149	168,405	174,115	185,295	0
SP5	185,286	178,176	174,870	172,232	174,141	168,396	174,110	185,286	0
SP6	185,301	178,186	174,881	172,243	174,152	168,410	174,117	185,301	0
SP7	185,314	178,197	174,892	172,255	174,166	168,422	174,127	185,314	0
SP8	185,296	178,182	174,874	172,237	174,150	168,404	174,116	185,296	0
SP9	185,284	178,175	174,869	172,231	174,140	168,395	174,109	185,284	0
RR1	185,285	178,180	174,874	172,234	174,145	168,396	174,110	185,285	0
RR2	185,308	178,190	174,886	172,248	174,158	168,416	174,125	185,308	0
RR3	185,272	178,164	174,858	172,223	174,129	168,382	174,096	185,272	0
RR4	185,287	178,178	174,870	172,231	174,143	168,397	174,110	185,287	0
RR5	185,284	178,172	174,868	172,229	174,137	168,394	174,107	185,284	0
RR6	185,291	178,181	174,874	172,236	174,147	168,401	174,115	185,291	0
RR7	185,285	178,177	174,871	172,232	174,141	168,398	174,109	185,285	0
RR8	185,265	178,158	174,851	172,216	174,130	168,382	174,093	185,265	0
RR9	185,289	178,179	174,873	172,234	174,144	168,399	174,112	185,289	0
SP2	185,298	178,187	174,883	172,241	174,151	168,407	174,120	185,298	0
SP3	185,301	178,189	174,882	172,247	174,158	168,413	174,123	185,301	0
SP4	185,291	178,180	174,873	172,235	174,145	168,397	174,110	185,291	0
SP5	185,279	178,168	174,859	172,224	174,135	168,387	174,101	185,279	0
SP6	185,263	178,159	174,844	172,207	174,113	168,367	174,081	185,263	0
SP7	185,262	178,157	174,847	172,210	174,119	168,369	174,095	185,262	0
SP8	185,301	178,189	174,881	172,243	174,155	168,413	174,120	185,301	0
SP9	185,293	178,182	174,876	172,236	174,147	168,405	174,115	185,293	0
T1 RRMS relative risk of progression	185,287	178,294	174,871	172,233	174,143	168,398	174,111	186,276	989
T1 RRMS relative risk of relapse	185,287	178,260	174,871	172,233	174,143	168,398	174,111	185,287	0
T2 RRMS relative risk of progression	185,287	178,178	174,728	172,233	174,143	168,398	174,111	186,235	947
T2 RRMS relative risk of relapse	185,287	178,178	174,852	172,233	174,143	168,398	174,111	185,287	0
T3 RRMS relative risk of progression	185,287	178,178	174,871	171,979	174,143	168,398	174,111	185,540	252
T3 RRMS relative risk of relapse	185,287	178,178	174,871	172,203	174,143	168,398	174,111	185,287	0

continued

TABLE 15 *Partial EVPI for individual parameters calculated using the one-level algorithm within the linear regression metamodel (all data in £) (cont'd)*

Parameter	ENB T1	ENB T2	ENB T3	ENB T4	ENB T5	ENB T6	ENB T0	Maximum ENB	EVPI for individual parameter
T4 RRMS relative risk of progression	185,287	178,178	174,871	172,233	174,270	168,398	174,111	186,100	813
T4 RRMS relative risk of relapse	185,287	178,178	174,871	172,233	174,180	168,398	174,111	185,287	0
T5 RRMS relative risk of progression	185,287	178,178	174,871	172,233	174,143	168,554	174,111	186,529	1,241
T5 RRMS relative risk of relapse	185,287	178,178	174,871	172,233	174,143	168,412	174,111	185,287	0
T6 RRMS relative risk of progression	185,287	178,178	174,871	172,233	174,143	168,398	174,297	186,157	870
T6 SPMS relative risk of progression	185,287	178,178	174,871	172,233	174,143	168,398	174,148	185,287	0
T6 RRMS relative risk of relapse	185,287	178,178	174,871	172,233	174,143	168,398	174,022	185,287	0
T6 SPMS relative risk of relapse	185,287	178,178	174,871	172,233	174,143	168,398	174,110	185,287	0

SE, standard error.

be included in the model. It was necessary to explore the relative importance of each of the parameters in the original ScHARR MS model, with the intention of ranking them according to their impact on the variability of the NBs. Due to the reasonably strong linear relationship between the sampled model inputs and the NBs, importance analysis was conducted using three regression-based factor screening methods:

1. standardised regression (β) coefficient analysis
2. PCV analysis
3. PRCC analysis.

As highlighted in Chapter 3, the first two of these techniques require the assumption of linearity between the model inputs and outputs. The PRCC method, however, does not assume that the relationship between model inputs and outputs is linear.

The three alternative methods were employed not only to highlight the most influential parameters within the model, but also to explore whether or not different techniques highlighted the same parameters as being important. The importance analysis was undertaken using 10,000 random samples generated from the original ScHARR MS model [see the section 'Is the model linear?' (p. 40)], and analysed using SPSS software. The analysis was undertaken for the linear model for T6 (Betaseron), as this included the relative risks of relapse and progression in secondary progressive health states. The methods are outlined below.

Standardised regression coefficient

The standardised regression coefficients were obtained from the regression analysis carried out in SPSS; the absolute values of coefficients were used since negative coefficients are equally influential on the ENBs as positive coefficients. The coefficients with the highest absolute values were considered the most important within the model. The accuracy of these coefficients is clearly dependent on the number of samples taken, although a data set containing 10,000 samples was considered sufficiently large to produce accurate results.

Partial contribution to variance

The PCV for each parameter was calculated by squaring the PCC of each parameter with the expected net benefit; hence the values ranged between 0 and 1.

Partial rank correlation coefficient

Unlike the other two methods of importance analysis described above, the PRCC method does not assume linearity between the model inputs

TABLE 16 *Comparison of parameter importance using three alternative methods*

	Standard regression coefficient		PCV method		PRCC method	
Rank	**Parameter**	**Absolute coefficient**	**Parameter**	**Squared coefficient**	**Parameter**	**Absolute PRCC**
1	Utility parameter *A*	0.7782	Utility parameter *A*	0.8953	Utility parameter *A*	0.9233
2	Cost of relapse	0.3737	Cost of relapse	0.6637	Utility parameter β	0.598
3	Cost parameter *H*	0.2140	Cost parameter *H*	0.3938	Cost of relapse	0.5318
4	Utility parameter β	0.2111	Utility parameter β	0.3849	Cost parameter *R*	0.491
5	Cost parameter *R*	0.1993	Cost parameter *R*	0.3593	Utility parameter α	0.4641
6	Utility parameter α	0.1799	Utility parameter α	0.3126	Cost parameter *H*	0.4523
7	RRMS relative risk of progression	0.1181	RRMS relative risk of progression	0.1653	RRMS relative risk of progression	0.3091
8	Cost parameter *L*	0.0430	Cost parameter *L*	0.0253	Year 1, 2 dropouts	0.1305
9	Year 1, 2 dropouts	0.0409	Year 1, 2 dropouts	0.0229	Cost parameter *L*	0.1128
10	SPMS relative risk of progression	0.0288	SPMS relative risk of progression	0.0119	Disutility of relapse	0.0865
11	Subsequent dropouts	0.0285	Subsequent dropouts	0.0116	SPMS relative risk of progression	0.0811
12	Disutility of relapse	0.0276	Disutility of relapse	0.0107	Subsequent dropouts	0.0791
13	RRMS mean sojourn state 0	0.0266	RRMS mean sojourn state 0	0.0099	RRMS mean sojourn state 0	0.0664
14	RRMS relative risk of relapse	0.0202	RRMS relative risk of relapse	0.0057	RRMS relapse count state 2	0.0651
15	RRMS relapse count state 2	0.0170	RRMS relapse count state 2	0.0040	RRMS relative risk of relapse	0.0513
16	RRMS relapse count state 1.5	0.0149	RRMS relapse count state 1.5	0.0032	RRMS relapse count state 2.5	0.0441
17	Relapse duration (days)	0.0136	RRMS relapse count state 0	0.0027	RRMS mean sojourn state 1	0.0437
18	RRMS relapse count state 0	0.0135	RRMS mean sojourn state 1	0.0025	Relapse duration (days)	0.0411
19	RRMS mean sojourn state 1	0.0133	Relapse duration (days)	0.0025	RRMS mean sojourn state 2	0.0358
20	RRMS relapse count state 2.5	0.0124	RRMS relapse count state 2.5	0.0023	RRMS relapse count state 1.5	0.0313
21	RRMS relapse count state 1	0.0111	RRMS relapse count state 1	0.0017	RRMS relapse count state 0	0.0261
22	RRMS mean sojourn state 2	0.0104	RRMS mean sojourn state 2	0.0015	RRMS mean sojourn state 9	0.0259
23	SPMS mean sojourn state 5	0.0096	SPMS mean sojourn state 5	0.0012	SPMS mean sojourn state 5	0.0253
24	RRMS mean sojourn state 4	0.0079	RRMS mean sojourn state 4	0.0009	RRMS relapse count state 9	0.0191
25	RRMS relapse count state 9	0.0074	RRMS relapse count state 9	0.0008	RRMS mean sojourn state 5	0.0144
26	SPMS mean sojourn state 7	0.0066	Side-effects utility adjustment	0.0006	SPMS relapse count state 6	0.0137
27	Side-effects utility adjustment	0.0063	SPMS mean sojourn state 7	0.0005	RRMS relapse count state 1	0.0133
28	RRMS mean sojourn state 9	0.0058	RRMS mean sojourn state 9	0.0005	Side-effects utility adjustment	0.0132
29	SPMS relapse count state 6	0.0057	SPMS relapse count state 6	0.0004	RRMS mean sojourn state 4	0.0123
30	RRMS mean sojourn state 5	0.0056	RRMS mean sojourn state 5	0.0004	SPMS mean sojourn state 7	0.0115

and outputs.[68] The PRCC of each parameter was calculated by deriving the rank correlation between each parameter and the NB and adjusting it for the effects of all other parameters in the model.

The results of the three importance analysis techniques are shown in *Table 16*.

Summary of importance analysis

The importance analysis showed that the 30 parameters selected by the three methods as inputs to the GP were identical. There were some slight discrepancies between the methods in terms of the order of the importance of these parameters, but in all three cases, several common variables were considerably more influential than the remaining variables. Primarily, these included the cost and utility parameters and the relative risks of disease progression associated with therapy; these were known to be key parameters within the ScHARR MS model. A central question arising from this analysis is 'how important does a parameter need to be for it to be considered influential?' There are no prescriptive rules determining this, hence the analyst's detailed knowledge of the model and of the behaviour of model parameters is key.

Undertaking EVPI analysis using a Gaussian process metamodel

We considered the use of a GP metamodel in obtaining estimates of the individual parameter partial EVPIs. An immediate difficulty faced was that the ScHARR MS model had 128 input parameters. We do not currently have software capable of constructing GP metamodels to high (input) dimensional models, and so we first selected what was believed to be the 30 most important input variables based upon the results of the importance analysis (see the section 'Factor screening to identify importance variables in the model' (p. 46)]. All other input variables were held constant at their mean values, hence in effect the GP metamodel was a 30-parameter simulation model.

A total of 300 sets of input parameter values were then chosen for each of the seven treatment options, designed to cover the entire sample space for the 30 input parameters. The ScHARR MS model was then run at these 2100 design points and GP metamodels were derived for the expected net benefits for the seven treatment options.

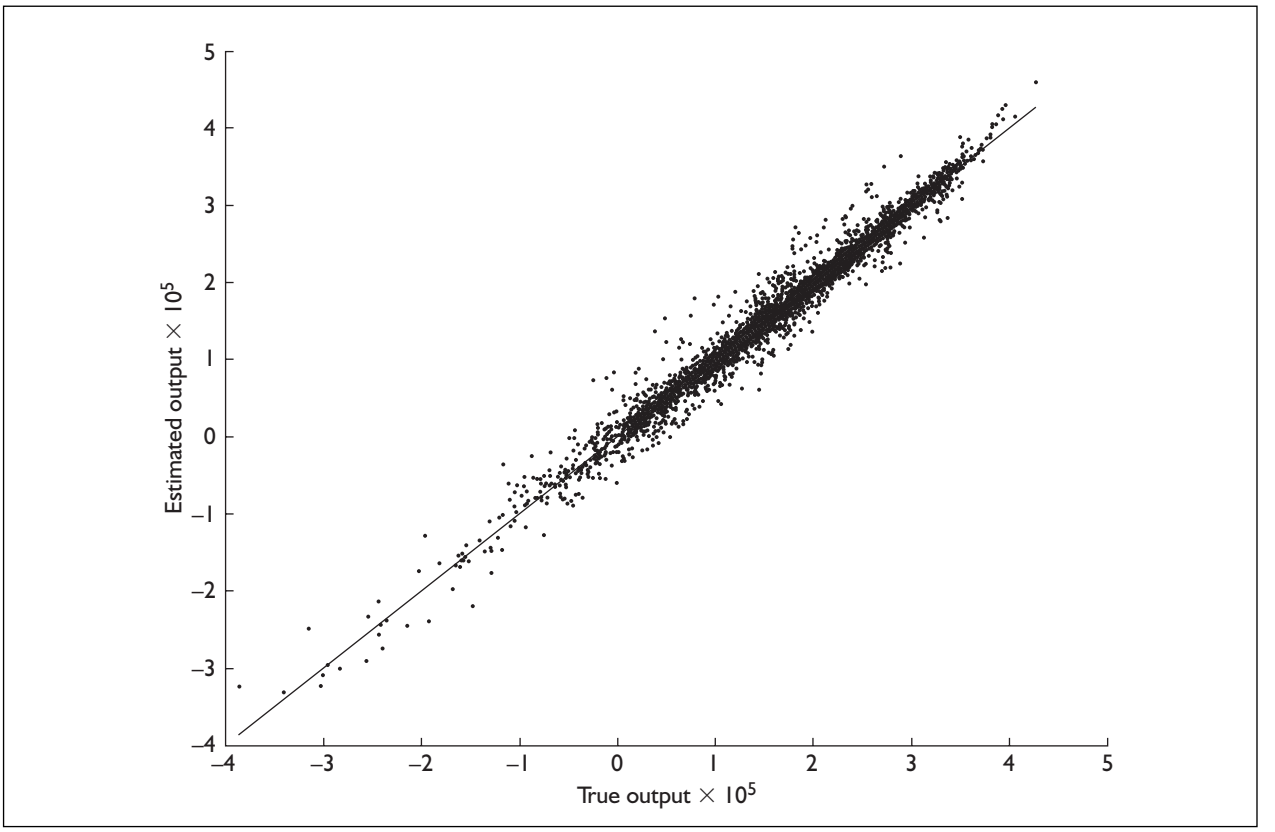

FIGURE 6 *True outputs versus predicted outputs using MS GP metamodel*

TABLE 17 *Mean absolute prediction errors using linear model and GP metamodel*

	Mean prediction errors for individual treatment strategies						
Model	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T0**
GP	9,569	9,351	9,428	8,781	10,079	11,073	9,186
Linear	18,852	19,421	19,984	19,445	20,204	19,589	20,259

To validate the GP metamodels, an additional 700 runs of the ScHARR MS model were generated and we then compared the true outputs for the seven treatment options with the predicted outputs from the seven metamodels. The results are shown in *Figure 6*.

The GP emulators were noticeably better at predicting the true outputs than the linear regression models. *Table 17* shows that the mean absolute prediction errors for the seven treatment strategies were approximately halved when using GP emulators.

The mean absolute errors from the GP emulators were between 1 and 2% of the observed ranges (largest observation minus the smallest observation) of the ENBs. Once the seven metamodels had been fitted, the global and partial EVPIs of the 30 parameters were estimated using a combination of Simpson's rule and Bayesian quadrature.[66]

The global EVPI was estimated to be £8771 per patient.

Table 18 clearly shows that the partial EVPI estimates generated using the GP metamodel are considerably higher than those generated using the ScHARR MS model and the regression metamodel via the one-level algorithm. This is likely to be due to the inclusion of all possible interactions between model parameters.

TABLE 18 *Partial EVPI results for individual parameters estimated using the GP metamodel*

Parameter	Partial EVPI (£)
Utility parameter *A*	215.50
Cost of relapse	595.90
Cost parameter *H*	1024.60
Utility parameter β	518.60
Cost parameter *R*	0
Utility parameter α	0
Dropouts for years 1, 2	3.80
Dropouts for subsequent years	8.30
Cost parameter *L*	0
T6 SPMS relative risk of progression	0
Disutility relapse	0
Mean sojourn time state 0 (RR)	0
Relapse count state 1.5 (RR)	0
Relapse duration (days)	0
Mean sojourn time state 1	0
Relapse count state 0 (RR)	1.20
Relapse count state 1 (RR)	0
Relapse count state 2.5 (RR)	0
Relapse count state 2 (RR)	0
Side-effect utility adjustment	0
Beta sojourn time state 8 (RR)	2.10
Mean sojourn time state 2 (RR)	0
Mean sojourn time state 4 (RR)	0
Beta sojourn time state 3 (RR)	0
Beta sojourn time state 4.5 (RR)	0
Relapse count state 3 (RR)	0
DSS EDSS ratio state 4 (RR)	9.50
Relapse count state 6.5 (SP)	0
T1 RRMS relative risk of progression	280.60
T2 RRMS relative risk of progression	626.00
T3 RRMS relative risk of progression	20.60
T4 RRMS relative risk of progression	1446.00
T5 RRMS relative risk of progression	1363.20
T6 RRMS relative risk of progression	258.70
T1 RRMS relative risk of relapse	0
T2 RRMS relative risk of relapse	0
T3 RRMS relative risk of relapse	0
T4 RRMS relative risk of relapse	0
T5 RRMS relative risk of relapse	0
T6 RRMS relative risk of relapse	0

Practical critique of the Gaussian process metamodel

Does it save time?

The technique is itself computationally intensive. The process of fitting the GP metamodel to the 300 ScHARR MS model runs for each treatment took several hours. Obtaining single-parameter partial EVPIs also took several hours to ensure numerical convergence of the estimates. Partial EVPIs for parameter groups were not considered here. In principle this could be done, at the expense of having to use Monte Carlo rather than numerical integration to evaluate the outer expectation. Clearly, the more computationally expensive the simulation model is, the more time-saving this technique will be.

Are there parametric restrictions for this method?
A significant drawback was the inability to handle all 128 parameters; importance analysis was required to reduce the number of parameter inputs for the GP metamodel to 30. There is therefore a possibility that the 30 selected parameters were not the 30 with the highest partial EVPIs. It may be possible to use this method with all 128 parameters and a larger number of model runs, although this would require programming ability in a fast language.

Closeness of results
With sufficient model runs using the ScHARR MS model, the GP metamodel would replicate the original cost-effectiveness model even more accurately and provide the correct partial EVPIs. This, however, would require many more than the 300 runs per treatment. Methods used by Oakley and O'Hagan[73] could be extended to derive the uncertainty in the partial EVPIs due to the use of the metamodel, although this approach itself involves some intensive computational effort. We do expect these estimates to be more accurate than those derived through the linear approximation, as the GP metamodel outperforms the linear model in predicting the true MS model output.

Ease of implementation
The implementation of the GP metamodel was the most complex of the methods considered within this case study. Implementation involves matrix algebra and numerical optimisation routines. Our software was written in MATLAB. A viable alternative to MATLAB would be the freely available R, but we believe that the computational requirements are beyond the capabilities of EXCEL.

Is the method algorithmic or heuristic?
GP metamodelling requires some heuristics application. It is commonly used to approximate complex computer simulation models but requires considerable specialist programming expertise.

Summary and conclusions

This chapter has used three different methods for estimating the EVPI for the case study, the results of which are compared below. The estimate of per patient global EVPI calculated using the original MS model was £8855, whereas the linear regression metamodel gave a slightly lower estimate of £7263. This discrepancy can be attributed to errors in the extreme values of the NBs using the regression model; although the mean NBs for each treatment strategy were very close to those generated using the original MS model, the errors arose in the calculation of the maximum NBs. It is likely that these biases arose owing to the imperfect linear relationship between the sampled model inputs and the associated net benefits, and also the loss of interactions between parameters within the linear regression metamodel. The GP metamodel estimated the global EVPI to be £8771; this was considerably more accurate than the linear approximation. The global EVPI estimate calculated using the original MS model should be considered to be more reliable than the EVPI calculating using the regression metamodel.

Clearly, it was not possible to estimate the partial EVPI for each parameter, or indeed groups of parameters, using the two-level EVPI algorithm alongside the ScHARR MS model itself owing to vast computational expense (see *Table 1*). As a result, two alternative metamodelling methods were used to replace the original cost-effectiveness model: linear regression metamodelling and GP metamodelling. However, the comparison of the global EVPIs calculated using the original MS model and the regression metamodel suggested a considerable degree of imprecision in the predictive ability of the regression model, as reflected by the lower estimate of global EVPI. The GP metamodel, however, gave a highly accurate estimate of the global EVPI, and the mean prediction errors from the observed data were considerably lower than those resulting from the linear regression metamodel (see *Table 17*). This suggests that the partial EVPIs estimated using the GP metamodel are an accurate approximation of the true EVPIs.

It can be seen from the regression metamodel that the sum of the partial EVPIs is almost equal to the global estimate calculated using the same method (~£7500), which highlights the fact that interactions between model parameters are not captured by the linear regression metamodel. The sum of the partial EVPI estimates using the GP metamodel was ~£6300; there is potentially some value in those parameters excluded from the GP simulation model. We would expect the Gaussian estimates to be more reliable owing to a better model fit, as demonstrated by the mean absolute prediction errors (see *Table 17*).

The GP metamodel is more reliable for this particular case study, as it is a non-linear method

that can deal with interactions between model variables. However, there exists an intrinsic problem here: due to the computational time required to perform a comprehensive EVPI analysis using the original ScHARR MS model, we have no basis for comparison of the Gaussian method. Whilst the GP methodology has been validated in a simpler case study,[66] this issue should be addressed in further case study policy problems.

The practical critique of the alternative metamodelling methods employed within the case study problem highlighted a number of advantages and disadvantages of the Gaussian methodology. In particular, the complexity of the technique, the absence of user-friendly modelling software and the limited number of input variables that can be included in the analysis mean that the method is readily accessible only to those with specialist expertise. However, most of these problems can be resolved through further research, and the benefits in terms of computational savings may be worthwhile, particularly for non-linear models that require considerable computational expense. Under certain conditions, such as instances whereby the relationship between sampled model inputs and NBs is reasonably linear, the linear regression approach may be adequate to highlight general areas for investment in further research.

Chapter 5

The expected value of perfect information for interferon-β and glatiramer acetate in the management of multiple sclerosis

Overview of case study results

This chapter summarises the results and conclusions of the EVPI analysis undertaken for the case study model of IFN-β and glatiramer acetate in the management of MS. The global EVPI estimate was calculated using the original ScHARR MS model, whereas the partial EVPI for individual parameters were calculated using the GP metamodel, as the Gaussian approximation is not biased by the imperfect linear relationship between the parameter inputs and net benefits [see the section 'Is the model linear?' (p. 40), or the loss of interactions between parameters.

Defining the relevant population for the decision

The MS population in England and Wales who would be potentially eligible for and might take up DMTs is itself subject to considerable uncertainty. In order to obtain an estimate of the population EVPI, it is necessary to scale up the previously calculated per patient EVPI by the population for whom the commissioning decision is relevant. The data used in estimating the relevant population for this decision are shown in *Table 19*.

EVPI results for interferon-β and glatiramer acetate in the management of MS

Global EVPI results

Table 20 shows the global EVPI estimates for each treatment strategy versus conventional management (T0), discounted over 10 years together with the decision EVPI across all parameters for all seven treatment strategies. These results were calculated using the original ScHARR MS model. As discussed in the section 'Correlation between treatment efficacies' (p. 13), the overall EVPI results are presented for two scenarios relating to different assumptions regarding the correlation between treatment efficacies. Assuming independent treatment effects the per patient EVPI results shown in *Table 20* suggests that the value of obtaining perfect information for all uncertain parameters within the case study model is £8855. This results in a population EVPI of £86,208,936 across the relevant MS population; this represents the upper

TABLE 19 *Data used for estimating the relevant population*

Variable	Value	Source
Prevalence	5700	Estimated using data collected from the 64 centres participating in the MS RSS Monitoring Study[74]
Incidence	3.8 per 100,000 population	Richards *et al.*[25]
% RRMS at onset	80%	NICE, *Multiple Sclerosis – national guidelines for NHS Management in Primary and Secondary Care*[75]
Percentage of patients eligible for DMTs	40–50%	Proportion of newly diagnosed RRMS who progress to DMTs (Boggild M, Walton Centre for Neurology and Neurosurgery: personal communication, 2003)
Annual incidence of DMT use	800	Calculated from incidence × population × %RRMS × Uptake
Discount rate	3.5%	NICE, *Guide to methods of technology appraisal* – Consultation document[76]

TABLE 20 *Global EVPI results for the ScHARR MS model (EVPI for individual treatment strategies and decision EVPI)*

	EVPI T1 vs T0	EVPI T2 vs T0	EVPI T3 vs T0	EVPI T4 vs T0	EVPI T5 vs T0	EVPI T6 vs T0	Decision EVPI
Per patient EVPI (£)	4,271	3,035	2,827	2,776	2,444	3,514	8,855
Population EVPI (£)	41,581,273	29,545,388	27,525,355	27,023,241	23,794,182	34,211,482	86,208,936

TABLE 21 *Two-level partial EVPI for parameters calculated using the GP metamodel (all data in £)*

Parameter	Per patient EVPI for individual parameter (£)	Population EVPI for individual parameter (discounted) (£)
T4 RRMS relative risk of progression	1,446.00	14,078,471.75
T5 RRMS relative risk of progression	1,363.20	13,272,318.60
Cost parameter *H*	1,024.60	9,975,658.48
T2 RRMS relative risk of progression	626.00	6,094,829.40
Cost of relapse	595.90	5,801,771.31
Utility parameter β	518.60	5,049,166.98
T1 RRMS relative risk of progression	280.60	2,731,963.47
T6 RRMS relative risk of progression	258.70	2,518,741.80
Utility parameter *A*	215.50	2,098,140.15
T3 RRMS relative risk of progression	20.60	200,564.67
DSS EDSS ratio state 4 (RRMS)	9.50	92,493.42
Subsequent drop outs	8.30	80,810.004
Year 1,2 drop outs	3.80	36,997.37
Beta sojourn time state 8 (RRMS)	2.10	20,445.91
Relapse count state 0 (RRMS)	1.20	11,683.38
Side effects utility adjustment	0.70	6,815.30
Relapse count state 1.5 (RRMS)	0.40	3,894.46
Utility parameter α	0.10	973.61
Cost parameter *R*	0.00	0.00
Cost parameter *L*	0.00	0.00
T6 SPMS relative risk of progression	0.00	0.00
Disutility relapse	0.00	0.00
Mean sojourn time state 0 (RRMS)	0.00	0.00
Relapse duration (days)	0.00	0.00
Mean sojourn time state 1	0.00	0.00
Relapse count state 1 (RRMS)	0.00	0.00
Relapse count state 2.5 (RRMS)	0.00	0.00
Relapse count state 2 (RRMS)	0.00	0.00
Mean sojourn time state 2 (RRMS)	0.00	0.00
Mean sojourn time state 4 (RRMS)	0.00	0.00
Beta sojourn time state 3 (RRMS)	0.00	0.00
Beta sojourn time state 4.5 (RRMS)	0.00	0.00
Relapse count state 3 (RRMS)	0.00	0.00
Relapse count state 6.5 (SPMS)	0.00	0.00
T1 RRMS relative risk of relapse	0.00	0.00
T2 RRMS relative risk of relapse	0.00	0.00
T3 RRMS relative risk of relapse	0.00	0.00
T4 RRMS relative risk of relapse	0.00	0.00
T5 RRMS relative risk of relapse	0.00	0.00
T6 RRMS relative risk of relapse	0.00	0.00

estimate for the overall EVPI. The global population EVPI estimate for strategy T1 versus T0 is £41,581,273; this represents the value of information assuming that treatment efficacies are perfectly correlated and hence represents a lower estimate for the overall EVPI.

Partial EVPI results for individual parameters calculated using the Gaussian process metamodel

Table 21 shows the estimates of partial EVPI for those parameters included in the GP metamodel.

Conclusions

This EVPI analysis of the MS cost-effectiveness model clearly suggests that large uncertainties surround many of the model parameters and that further research is merited on the impact of IFN-β and glatiramer acetate. Specifically, this research should focus on the relationship between the EDSS and the cost of care, the relationship between the EDSS and QoL, the rate at which patients drop off therapy and, in particular, the impact of these therapies on disease progression.

The primary focus of this analysis has been on estimating the 'per patient' value of information; however, the population scaling factors are also highly uncertain, and especially the eligibility for and uptake of DMTs is particularly uncertain, being currently based upon subjective judgement. This is an area for further research.

The large difference in the EVPIs obtained from the completely correlated and independent model assessments indicates that further knowledge on the correlation between treatment efficacies would be highly valuable in commissioning decision-making. This is intuitively sensible as learning something about the treatment efficacy of one of the DMTs is likely to give information about other drugs in this set. Furthermore, the EVPI assessment incorporating all treatment options assumes that a single treatment is identified as optimal and selected for commissioning on the basis of its maximum net benefit. Given the high degree of uncertainty and the likely small differences in treatment efficacy and consequential net benefit, an exclusive commissioning recommendation identifying one specific product is unlikely and a broad commissioning decision covering groups of or all products is more probable. Given this, the lower estimate for the expected value of information is likely to be a better representation of the value of further research than the upper estimate.

Although further information on costs associated with particular EDSS states and the rates at which patients drop off therapy may be obtained through non-experimental designs such as observational studies, further useful information on the impact of DMTs on disease progression and associated health outcomes would be most reliably obtained through a long-term RCT which includes a direct assessment of QoL.

Chapter 6

Discussion and conclusions

Process for undertaking EVPI analysis

The following outlines a general framework undertaking EVPI analysis within a decision analytic economic model. This framework is presented as a sequential approach to methods available for undertaking EVPI analysis; however, many unresolved methodological issues exist, hence it should not be considered as a prescriptive algorithm.

1. Estimate the global EVPI using a Monte Carlo simulation [see the section 'Calculating the overall EVPI across all parameters simultaneously' (p. 5)]
2. Estimate the processing time required to calculate partial EVPIs using the full two-level Monte Carlo sampling algorithm [see the sections 'Can EVPI be calculated numerically?' (p. 17) and 'Can EVI be calculated numerically?' (p. 39)].
3. If it is feasible to use the two-level Monte Carlo method to generate a full set of partial EVPIs, then use the algorithm described in *Box 2* (p. 5)
4. If it is not feasible to undertake the full two-level Monte Carlo analysis, test the linearity of the model. If the whole model is linear (or if the model is linear in the required subsets of parameters), use the one-level Monte Carlo algorithm outlined in *Box 3* (p. 7).
5. If the full analysis is not feasible **and** if the model is non-linear, construct a metamodel to approximate the underlying health economic model, using importance analysis techniques to identify a subset of key parameters for use in the metamodel.

   Given the current state of knowledge, it is not possible to set down criteria to govern the choice of metamodelling methodology. However, in general the simpler and more accessible methodologies are open to greater predictive error, whereas the more complex methods may be more accurate but considerably more difficult to implement, particularly in the absence of specialist expertise and/or dedicated software support.
6. Use the metamodel to calculate the partial EVPIs for parameters.

Methodological issues for consideration

Linearity of the model

This report has highlighted the central role of regression analysis in performing EVPI analysis. The main potential drawback concerns the degree of linearity between the model inputs and the NB. If the relationship between NBs and the parameter inputs is weak, linear regression metamodelling is unlikely to be useful in performing partial EVPI analysis. Conversely, if the relationship is strongly linear, that is, an adjusted r^2 value which is close to 1, it is likely that even if the ENBs for each treatment strategy are predicted with accuracy, the prediction error in the calculation of net benefits is likely to be magnified in the calculation of EVPI. The applied methodology presented in Chapter 4 clearly points towards using more complex and sophisticated metamodelling approaches in order to obtain greater accuracy in EVPI estimation.

The regression metamodel may be used in order to obtain one-level estimates of partial EVPI for all model parameters. Although the estimates of partial EVPI for parameters obtained using this method are unlikely to be perfectly accurate, the exercise may enable the modeller to ascertain which of the individual model parameters are likely to attain value and which are not, and potentially suggest an order of magnitude for this expected value. The fundamental benefit of this deductive approach is that if the analyst is aware of the key parameters, it may be possible to revert back to the original cost-effectiveness model and perform partial EVPI analysis using the correct two-level sampling algorithm on those identified parameters and to ignore the remaining parameter set. This will depend, however, on the time required to run the analysis using the two-level algorithm on a single model parameter and on the specified number of parameters of interest.

Although the question of 'how linear is linear enough?' for use in EVPI analysis cannot be resolved using standard statistical tests, it is possible to explore the degree of approximation error resulting from a linear regression metamodel

through comparing the global EVPI results calculated using the regression metamodel and the global EVPI results calculated using the original cost-effectiveness model. If the two global EVPI results are similar (and the adjusted r^2 value suggests a strong linear relationship), this should enable the analyst to gauge the degree to which non-linearity may distort the results of the partial EVPI analysis. If there is a considerable error between the global EVPI estimates, this should forewarn against using the one-level EVPI algorithm, and highlight the need for more sophisticated non-linear methods such as GP metamodelling.

The importance analysis techniques required to identify those parameters which have the greatest impact upon NB are largely reliant on linear regression analysis; it is possible that if the model itself is highly non-linear, this process may result in the identification of unimportant variables and the omission of important variables. In such cases, non-linear factor screening methods such as the partial rank correlation technique should be used.

Summary of use of metamodelling techniques

Chapter 3 presented some of the main metamodelling techniques available and attempted to describe the suitability of each method for different scenarios. It should be noted that although metamodels allow faster analysis of a problem, their use introduces an added element of uncertainty to the analysis; a metamodel can only ever approximate a system rather than fully replace it. Although many of the techniques appear similar in theory (i.e. many of the techniques are based on the principles of regression), the main difference relevant to the users of health economic models concerns the ease of use and availability of software. Many of these techniques have been applied in only a limited number of case studies, hence their suitability for use within EVPI analysis has not been demonstrated.

The suitability of these alternative metamodelling methods in performing EVPI analysis will essentially be determined by the expertise of the modeller, the time available for the project and the degree of accuracy required in the results.

It is not unreasonable to postulate that when faced with a computationally expensive decision model, the general user of health economic models is primarily concerned with selecting the easiest and quickest metamodelling technique which provides reasonably accurate results. Indeed, in instances whereby the original cost-effectiveness model is approximately linear, regression metamodelling may be an adequate approach for identifying areas for investment in further research. This review has identified several classes of metamodelling technique. Although it has been possible to identify some of their more generic characteristics, these are certainly insufficient to identify one generally preferred technique, but also insufficient to identify a set of criteria for selecting a specific technique given specific case study characteristics.

Limitations of this study

This review has investigated the use of alternative metamodelling methods using a computationally expensive case study model.

1. The information currently available in the public domain on the alternative metamodelling techniques is limited. Hence insufficient information was available concerning the practical application of several of the metamodelling methodologies reviewed in Chapter 3. These methods could not be confidently applied to the case study model.
2. The complexity of the ScHARR MS model means that it is infeasible to generate the partial EVPI analysis using the two-level sampling algorithm, hence this means that there is no direct means of validating fully the partial EVPIs calculated using either the one-level sampling algorithm, the linear regression metamodel or the GP metamodel. Direct tests of validity have only been possible on the estimate of overall EVPI.
3. It has been demonstrated during the course of this work that there is a high degree of linearity between sampled parameter values and net benefits generated by the ScHARR MS model. This means that the exploration of the impact of non-linearities on the predictive ability of the metamodels considered and of the impact on parameter selection via importance analysis has been limited.

Further research

This review has highlighted a number of areas requiring further research:

- *Further research indicated by the case study*. The partial EVPI estimates generated using both the linear regression metamodel and the GP

metamodel suggest that further research concerning the relationship between the EDSS, costs of care and health outcomes, the rates at which patients drop off therapy and in particular the impact of DMTs on the progression of MS is required.

- *Inclusion of the 'relevant population' within the sensitivity analysis*. Previous value of information studies have calculated the population EVPI by simply multiplying the per patient EVPI by a fixed number of patients over the lifetime of the decision. However, as the population relevant to a particular decision is itself uncertain, there remains an unresolved methodological issue concerning whether the uncertainty in the epidemiological parameters should also be accounted for within the sensitivity analysis.
- *Development of criteria for selecting a metamodelling approach*. Any health economic model could be replaced by a metamodel. There exist a number of other such techniques which have not been presented in this review which are currently not widely used and are still the subject of ongoing research. Methodological and case study work would be of benefit in exploring the application of these metamodelling techniques within health economic models and in the specific application to EVPI analysis. Of key interest would be to investigate whether there are any characteristics of EVPI analyses that make specific metamodelling techniques more preferable; for example, a good approximation is only required within a restricted domain of the sample space, that is, where a commissioning decision changes. Comparative assessments using the different techniques applied to common case studies would also be beneficial in informing ease of use, the level of expertise required and accuracy of results for each metamodelling technique.
- *The use of metamodelling for EVSI and ENBS analysis*. The role of metamodelling techniques in EVSI and ENBS requires further research. Due to similarities in the algorithms used, it is reasonable to suggest that metamodelling could have an instrumental role in performing EVSI and ENBS analysis for computationally expensive models.
- *Use of alternative decision analytic software*. The framework has assumed that the general user of health economic models uses EXCEL as the primary decision tool. Clearly, there is the possibility of using other software packages to build such models, which may increase their efficiency and thus, in certain cases, obviate the need to resort to metamodelling techniques.

Acknowledgements

We would like to acknowledge the support of Mr Alan Brennan, Director of Operational Research at ScHARR, and Professor Tony O'Hagan, Director of the Centre for Bayesian Statistics in Health Economics, for their support of this work. We would also like to thank Ms Meenal Patwardhan and colleagues for making their review of published EDSS cost studies available to the authors. We would like to acknowledge the invaluable input of Mr Stephen Palmer, Dr Allan Wailloo and Dr David Worthington, who peer reviewed this report.

Contributions of authors

Paul Tappenden (Research Fellow in Operational Research), Jim Chilcott (Senior Research Fellow in Operational Research) and Chris McCabe (Senior lecturer in Health Economics) developed the 'public' ScHARR cost-effectiveness model. Simon Eggington (Research Associate in Operational Research) conducted the systematic review of methods for importance analysis and metamodelling and undertook the importance analysis for the ScHARR model. Paul Tappenden, Jim Chilcott and Simon Eggington developed the methodological framework upon which this review is focused, and undertook the value of information analysis for the case study model. Paul Tappenden constructed the linear regression metamodel for the case study. Jeremy Oakley (Lecturer in Probability and Statistics) constructed the Gaussian process metamodel for the case study.

About ScHARR

The School of Health and Related Research (ScHARR) is one of the four Schools that comprise the Faculty of Medicine at the University of Sheffield. ScHARR brings together a wide range of medical- and health-related disciplines including public health, general practice, mental health, epidemiology, health economics, management sciences, medical statistics, operational research and information science. It includes the Sheffield unit of the Trent Institute for Health Services Research, which is funded by NHS R&D to facilitate high-quality health services research and capacity development.

The ScHARR Technology Assessment Group (ScHARR-TAG) synthesises research on the effectiveness and cost-effectiveness of healthcare interventions for the NHS R&D Health Technology Assessment Programme on behalf of a range of policy makers, including the National Institute of Clinical Excellence. ScHARR-TAG is part of a wider collaboration of six units from other regions. The other units are: Southampton Health Technology Assessment Centre (SHTAC), University of Southampton; Aberdeen Health Technology Assessment Group (Aberdeen HTA Group), University of Aberdeen; Liverpool Reviews & Implementation Group (LRiG), University of Liverpool; Peninsular Technology Assessment Group (PenTAG), University of Exeter; NHS Centre for Reviews and Dissemination, University of York; and West Midlands Health Technology Assessment Collaboration (WMHTAC), University of Birmingham.

References

1. Brennan A, Chilcott JB, Kharroubi A, O'Hagan A. Calculating expected value of perfect information: resolution of conflicting methods via a two level Monte Carlo approach. URL: http://www.shef.ac.uk/chebs/FFortnights/FF2/Brennan.doc. Accessed 17 February 2003.
2. Palmer S, Smith PC. Incorporating option values into the economic evaluation of health care technologies. *J Health Econ* 2003;**19**:755–66.
3. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 1999;**3**(2).
4. Raiffa H. *Decision Analysis: Introductory Lectures on Choice Under Uncertainty*. Reading, MA: Addison-Wesley; 1968.
5. Thompson MS. Decision-analytic determination of study size: the case of electronic fetal monitoring. *Med Decis Making* 1981;**1**:165–79.
6. Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. *Med Decis Making* 1988;**8**:279–89.
7. Eddy DM. Research Planning. In Eddy DM, Hasselblad V, Shachter RD, editors. *Meta-analysis by the confidence profile method. The statistical synthesis of evidence*. Boston, MA: Academic Press. 1992. pp. 313–22.
8. Hornberger JC, Brown BW, Halpern J. Designing a cost effective clinical trial. *Stat Med* 1995; **14**:2249–59.
9. Hornberger J, Eghtesady P. The cost benefit of a randomised trial to a health care organisation. *Control Clini Trials* 1998;**19**:198–211.
10. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ* 1996;**5**:513–24.
11. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;**18**:341–64.
12. Felli C, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Med Decis Making* 1998;**18**:95–109.
13. Felli JC, Hazen GB. A Bayesian approach to sensitivity analysis. *Health Econ* 1999;**8**:263–8.
14. Chilcott JB, Brennan AB, Booth A, Karnon J, Tappenden P. The role of modelling in prioritising and planning clinical trials. *Health Technol Assess* 2003;**7**(23).
15. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). *Lancet* 2002;**360**:711–15.
16. Fenwick E, Claxton K, Sculpher M, Briggs A. *Improving the efficiency and relevance of health technology assessment: the role of iterative decision analytic modelling*. CHE Discussion Paper 179. York: University of York, Centre for Health Economics. 2000.
17. Claxton K, Neumann PJ, Araki S, Weinstein MC. Bayesian value-of-information analysis. An application to a policy model of Alzheimer's disease. *Int J Technol Assess Health Care* 2001;**17**:38–55.
18. Payne N, Chilcott JB, McGoogan E. Liquid-based cytology in cervical screening: a rapid and systematic review. *Health Technol Assess* 2000;**4**(18).
19. Wight JP, Chilcott JB, Holmes MW, Brewer N. The clinical and cost-effectiveness of pulsatile machine perfusion versus cold storage of kidneys for transplantation retrieved from heart-beating and non-heart-beating donors. *Health Technol Assess* 2003;**7**(25).
20. Pandor A, Eastham J, Beverley C, Chilcott J, Paisley S. Clinical and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review. *Health Technol Assess* 2004;**8**(12).
21. Claxton K, Ginnelly L, Sculpher M, Palmer S, Philips ZMJ. Pilot study of using value of information analysis to set priorities in research and development. *Health Technol Assess* 2004 (in press).
22. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making* 2003;**22**:290–308.
23. Chilcott JB, McCabe CM, Tappenden P, O'Hagan A, Cooper NJ, Claxton K, *et al*. Modelling the cost effectiveness of interferon beta and glatiramer acetate in the management of multiple sclerosis. *BMJ* 2003;**326**:522.
24. Compston A, Ebers G, Lassman H, McDonald I, Matthews B, Wekerle H. *McAlpine's multiple sclerosis*. London: Churchill Livingstone; 1998.
25. Richards RG, Beard SM, Sampson FC, Tappenden P. A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models. *Health Technol Assess* 2002;**6**:1–73.

26. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale. *Neurology* 1983;**33**:1444–52.

27. Kobelt G, Lindgren P, Parkin D, Francis DA, Johnson M, Bates D, *et al*. *Costs and quality of life in multiple sclerosis. A cross-sectional observational study in the UK*. Scandinavian Working Papers in Economics; 2000; URL: http://www.swopec.hhs.se/hastef/papers/hastef0398.pdf. Accessed 12 May 2001.

28. Chadwick D, Compston A, Miller D, ZajicekJ, Dick J, Martyn C, *et al*. *Guidelines for the use of beta interferons and glatiramer acetate in multiple sclerosis*. London: Association of British Neurologists; 2001. URL: http://www.theabn.org/downloads/msdoc.pdf. Accessed 14 August 2003.

29. NHS Executive. EL 94 72. Leeds: Department of Health; 1994.

30. Brown MG, Murray TJ, Sketris IS, Fisk JD, LeBlanc JC, Schwartz CE, *et al*. Cost-effectiveness of interferon beta-1b in slowing multiple sclerosis disability progression. First estimates. *Int J Technol Assess Health Care* 2000;**16**:751–67.

31. Kendrick M, Johnson KI. Long-term treatment of multiple sclerosis with interferon-beta may be cost effective. [Review] [17 refs]. *Pharmacoeconomics* 2000;**18**:45–53.

32. Forbes RB, Lees A, Waugh N, Swingler RJ. Population based cost utility study of interferon beta-1b in secondary progressive multiple sclerosis [comment]. *BMJ* 1999;**319**:1529–33.

33. National Institute for Clinical Excellence. Minutes of the technology appraisals committee meeting 13th December 2000. URL: http://www.nice.org.uk/article.asp?a=16091. Accessed 6 May 2001.

34. National Institute for Clinical Excellence. *Full guidance on the use of beta interferon and glatiramer for the treatment of multiple sclerosis*. Technology Appraisal No.32. 2002.URL: http://www.nice.org.uk/Docref.asp?d=27634. Accessed 8 April 2003.

35. Tappenden P, Chilcott JB, O'Hagan A, McCabe C, Cooper NJ, Abrams KA, *et al*. *Cost-effectiveness of beta interferon and glatiramer acetate in the management of multiple sclerosis*. Report to the National Institute for Clinical Excellence. 2001.

36. Ebers, G. London Ontario cohort study: confidential data on file at National Institute for Clinical Excellence. London. 2001.

37. IFNB Multiple Sclerosis Study Group. Interferon beta-lb in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Am Acad Neurol* 1995;**45**:1277–85.

38. PRISMS Study Group. Randomized double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. *Lancet* 1998;**352**:1498–504.

39. Jacobs LD, Cookfair DL, Rudick RR, Herndon RM, Richet JR, Salazir AM. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Am Neurol Assoc* 1996;**39**:285–94.

40. Johnson KP, Brooks BR, Cohen JA, Ford CC , Goldstein J, Lisak RP. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multi centre, double-blind, placebo-controlled trial. *Am Acad Neurol* 1995;**45**:1268–76.

41. European Study Group on Interferon Beta-1b in Secondary Progressive MS. Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. *Lancet* 1998;**352**:1491–7.

42. Patwardhan M, Matchar D, Samsa D, McCrory D, Li T, Williams R. Cost of multiple sclerosis by level of disability: a review of literature. Poster presented 22 June 2003 at International Society for Technology Assessment in Health Care, Canmore, Alberta, 2003.

43. Murphy N, Confavreux C, Haas J, König N, Roullet E, Sailor M, *et al*. Economic evaluation of multiple sclerosis in the UK, Germany and France. *Pharmacoeconomics* 1998;**13**:607–22.

44. Parkin D, McNamee P, Jacoby A, Miller P, Thomas S, Bates D. A cost–utility analysis of beta interferon for multiple sclerosis. *Health Technol Assess* 1998;**2**(4).

45. Grima DT, Torrance GW, Francis G. Cost and health related quality of life consequences of multiple sclerosis. *Mult Scler* 2000;**6**:91–8.

46. Department of Health. Reference costs 2002. URL: http://www.doh.gov.uk/refcosts2002/collection/download.htm. Accessed 5 April 2003.

47. Barton RR. Simulation metamodels. In Medeiros DJ, Watson EF, Carson JS, Manivannan MS, editors. *Proceedings of the 1998 Winter Simulation Conference*. Washington. 1998.

48. Hurrion RD, Birgil S. A comparison of factorial and random experimental design methods for the development of regression and neural network simulation metamodels. *J Oper Res Soc* 1999; **50**:1018–33.

49. Hussain MF, Barton RR, Joshi SB. Metamodelling: radial basis functions versus polynomials. *Eur J Oper Res* 2002;**138**:142–54.

50. Batmaz I, Tunali S. Small response surface designs for metamodel estimation. *Eur J Oper Res* 2003; **145**:455–70.

51. Jin R, Chen W, Simpson TW. Comparative studies of metamodelling techniques under multiple

modelling criteria. *Struct Multidisc Optimis* 2001; **23**:1–13.

52. Kleijnen JPC, Sargent RG. A methodology for fitting and validating metamodels in simulation. *Eur J Oper Res* 2000;**120**,14–29.

53. Sasena MJ. Optimisation of computer simulations via smoothing splines and Kriging metamodels. Masters thesis, University of Michigan, 1998.

54. Meckesheimer M, Barton RR, Simpson TW, Limayem F, Yannou B. Metamodelling of combined discrete/continuous responses. *Am Inst Aeronaut Astronaut J* 2001;**39**:1950–9.

55. Montgomery DC. *Design and analysis of experiments*. New York: Wiley, 1991.

56. Hurrion RD. Using a neural network to enhance the decision making quality of a visual interactive simulation model. *J Oper Res Soc* 1992;**43**:333–41.

57. Kilmer RA, Smith AE, Shuman LJ. Computing confidence intervals for stochastic simulation using neural network metamodels. *Comp Ind Eng* 1999;**36**:391–407.

58. Marseguerra M, Masini R, Zio E. Neural modelling for sensitivity analysis in reliability and availability calculations. In *Proceedings of European Safety and Reliability International Conference*. 16–20 September 2001. Torino, Italy. pp. 189–196.

59. StatSoft. Multivariate adaptive regression splines (MARSplines). URL: www.statsoftinc.com/textbook/stmars.html. Accessed 23 July 2003.

60. Myers RH, Montgomery DC. *Response surface methodology: process and product optimization using designed experiments*. 2nd edition. New York: Wiley. 2002.

61. Carson Y, Maria A. Simulation optimisation: methods and applications. In Andradottir S, Healy KJ, Withers DH, Nelson BL, editors *Proceedings of the 1997 Winter Simulation Conference*. Atlanta. 1997. pp. 118–26.

62. Simpson TW, Mauery TM, Korte JJ, Mistree F. Comparison of response surface and Kriging models for multidisciplinary design optimization. In *Proceedings of the 7th Symposium on Multidisciplinary Analysis and Optimization, St Louis, MO*. AIAA-98-4755. 1998.

63. Koch PN, Mavris D, Mistree F. Multi-level, Partitioned Response Surfaces for Modelling Complex Systems. In *Proceedings of the 7th Symposium on Multidisciplinary Analysis and Optimisation, St Louis, MO*. AIAA-98-4958. 1998.

64. O'Hagan A. Curve fitting and optimal design for prediction (with discussion). *J R Stat Soc Ser B* 1978;**40**:1–42.

65. O'Hagan A, Kennedy MC, Oakley JE. Uncertainty analysis and other inference tools for complex computer codes (with discussion). In Bernado JM, editor *Bayesian statistics 6*. Oxford: Oxford University Press; 1999. pp 503–24.

66. Oakley, J. E. *Value of information for complex cost-effectiveness models*. Research Report No. 533/02. Sheffield: Department of Probability and Statistics, University of Sheffield; 2002

67. Simpson TW, Peplinski J, Koch PN, Allen JK. Metamodels for computer-based engineering design: survey and recommendations. *Eng Comput* 2001;**17**:129–50.

68. Coyle D, Buxton M, O'Brien B. Measures of importance for economic analysis based on decision modelling. *J Clin Epidemiol* 2003; **56**:989–97.

69. McKenna SA, Arnold BW. Quantitative sensitivity and importance analysis of parameters in Monte Carlo models. In *Proceedings of the International High-level Radioactive Waste Management Conference*. La Grange, IL: American Nuclear Society; 1998.

70. Bettonvil B, Kleijnen JPC. Searching for important factors in simulation models with many factors: sequential bifurcation. *Eur J Oper Res* 1996; **96**:180–94.

71. Morrice D, Schruben L. Simulation factor screening using harmonic analysis. *Manage Sci* 1993;**39**:1459–76.

72. Cheng RCH. Searching for important factors: sequential bifurcation under uncertainty. In Andradottir S, Healij KJ, Withes DH, Nelson BL, editors *Proceedings of the 1997 Winter Simulation Conference*. Atlanta. 1997. pp. 275–80.

73. Oakley JE, O'Hagan A. Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika* 2002;**89**:769–84.

74. School of Health and Related Research. *MS risk share scheme monitoring study: Appendix 1 – drug data*. November Report. Sheffield: ScHARR; 2003.

75. National Institute for Clinical Excellence. *Multiple sclerosis – national guidelines for NHS management of primary and secondary care (draft for 2nd round stakeholder consultation)*. London: NICE; 2003.

76. National Institute for Clinical Excellence. *Guide to methods of technology appraisal*. Consultation Document. London: NICE; 2003.

Appendix 1

The expanded disability status scale[26]

Status	Description
0.0	Normal neurological examination
1.0	No disability, minimal symptoms
1.5	No disability, minimal signs in more than one functional system
2.0	Slightly more disability in one functional system
2.5	Slightly greater disability in two functioning systems
3.0	Moderate disability in one functional system; fully ambulatory
3.5	Fully ambulatory but with moderate disability in one functional system and more than minimal disability in several others
4.0	Fully ambulatory without aid, self-sufficient, up and about ~ 12 hours per day despite relatively severe disability; able to walk without aid or rest ~ 500 m
4.5	Fully ambulatory without aid, up and about much of the day, able to work a full day, may otherwise have some limitation of full activity or require minimal assistance; characterised by relatively severe disability; able to walk without aid or rest ~ 300 m
5.0	Ambulatory without aid or rest for ~ 200 m; disability severe enough to impair full daily activities (work a full day without special provisions).
5.5	Ambulatory without aid or rest for ~ 100 m; disability severe enough to preclude full daily activities
6.0	Intermittent or unilateral constant assistance (cane, crutch, brace) required to walk ~ 100 m with or without resting
6.5	Constant bilateral assistance (canes, crutches, braces) required to walk ~ 20 m without resting
7.0	Unable to walk beyond ~ 5 m even with aid, essentially restricted to wheelchair; wheels self in standard wheelchair and transfers alone; up and about in wheelchair ~ 12 hours per day
7.5	Unable to take more than a few steps; restricted to wheelchair; may need aid in transfer; wheels self but cannot carry on in standard wheelchair a full day; may require motorised wheelchair
8.0	Essentially restricted to bed or chair or perambulated in wheelchair, but may be out of bed itself much of the day; retains many self-care functions; generally has effective use of arms
8.5	Essentially restricted to bed much of day; has some effective use of arms; retains some self-care functions
9.0	Helpless bed patient; can communicate and eat
9.5	Totally helpless bed patient; unable to communicate effectively or eat/swallow
10.0	Death due to MS

Appendix 2

Parameter group descriptions

Group no.	Group name	Description of parameters included in group
1	EDSS costs	These parameters describe the relationship between cost of care and EDSS state.
2	EDSS utilities	These parameters describe the relationship between utility and EDSS state.
3	Relapse – duration, cost and disutility	These refer to the duration of relapse and its associated cost. The disutility refers to the decrement applied to reflect the temporary reduction in an individual's QoL whilst experiencing relapse.
4	EDSS duration – beta sojourn times	These refer to the expected time to progress from an RRMS EDSS state to the equivalent SPMS health state
5	Relative risks of treatment effect	These are the treatment-specific parameters which describe the relative risk of progression and relapse for RRMS and SPMS
6	Side-effects	These parameters describe the proportion of the cohort experiencing treatment-related side-effects and their associated utility decrement
7	Dropouts	These parameters describe the proportion of the cohort who drop off therapy during the first 2 years of treatment and the proportion who drop off therapy during a subsequent model cycle
8	Relapse count	These parameters relate to the average number of relapses experienced in each EDSS state
9	Mean sojourn times in EDSS states	These parameters describe the mean duration a patient will stay in each EDSS state
10	DSS to EDSS ratios	These parameters are used to map the DSS sojourn times to the EDSS scale

Appendix 3

Sample size for Monte Carlo partial EVPI calculations

Suppose we have T treatment options and our economic model computes the net benefit $NB(t, \mathbf{x})$ for treatment $t = 1, \ldots, T$, when provided with input parameter x. We will denote the true, uncertain values of the input parameters by $\mathbf{X}$, so that the true, uncertain net benefit of treatment t is given by $NB(t, X)$. Writing $\mathbf{X} = \{X_1, \ldots, X_d\}$, the partial EVPI for the ith parameter X_i is given by

$$E_{x_i}\left[\max_t \; E_{X|x_i}\{NB(t, X)|X_i\}\right] - \max_t \; E_X\{NB(t, X)\} \tag{1}$$

The majority of the computational effort required to estimate the partial EVPIs of all the parameters results from having to estimate the first of these two terms for $i = 1, \ldots, d$. Monte Carlo sampling can be used to estimated this term. The Monte Carlo estimate is derived in two stages:

Stage 1:
$$\hat{E} = \hat{E}_{X_i}\left[\max_t \; E_{\mathbf{X}}\{NB(t, \mathbf{X})|X_i\}\right] = \frac{1}{N}\sum_{j=1}^{N} m(X_{i,j}) \tag{2}$$

with $X_{i,1}, \ldots, X_{i,\mathrm{N}}$ randomly sampled from the distribution of X_i and

$$m(X_i) = \max_t \; E_{\mathbf{X}}\{NB(t, \mathbf{X})|X_i\} \tag{3}$$

It will then be necessary to estimate $m(X_{i,j})$ by $\hat{m}(X_{i,j})$ using Monte Carlo:

Stage 2:
$$\hat{m}(X_{i,j}) = \max_t \frac{1}{M}\sum_{k=1}^{M} NB(t, \mathbf{X}_k) \tag{4}$$

with $\mathbf{X}_k$ randomly sampled from the distribution of $\{X_1, \ldots, X_d\}$ conditional on $X_\mathrm{i} = X_{i,j}$.

Overall, this will require a total of $N \times M \times d$ model evaluation (assuming that a single model evaluation gives the net benefit for all T treatments).

For the stage 1 approximation, if we are considering a one-dimensional variable X_i, then an alternative to Monte Carlo estimation is to use numerical integration such as Simpson's rule. However, it is important to appreciate that with Simpson's rule, if we use a small value of N (with $X_{i1}, \ldots, X_{i,N}$ chosen deterministically according to the quadrature rule), we will need a large value of M. Random noise in the second stage will not 'cancel itself out' in the first stage if N is small.

If Monte Carlo sampling is used, then for sufficiently large N, a 95% CI can be derived for the estimator of the partial EVPI $\hat{E}$ as

$$\{\bar{m}(X_i) - 1.96\hat{\sigma}(X_i), \bar{m}(X_i) + 1.96\hat{\sigma}(X_i)\} \tag{5}$$

with $\bar{m}(X_i)$ and $\hat{\sigma}^2(X_i)$ the sample mean and variance of $\hat{m}(X_{i,1}), \ldots \hat{m}(X_{i,N})$ respectively.

For N sufficiently large (using Monte Carlo or numerical integration), $\hat{E}$ will converge to a single value, but not necessarily the **correct** value of the partial EVPI. This will depend on the stage 2 approximation. For this approximation, Monte Carlo will generally be the only viable option owing to the dimensionality of $\mathbf{X}$. However, if M is too small, then $\hat{m}(X_{i,j})$ is likely to be a **biased** estimator of $m(X_{i,j})$. Using the notation

$$\mu_t(X_i) = E_{\mathbf{X}|x_i}\{NB(t, \mathbf{X})\}X_i) \tag{6}$$

$$U_t(X_i) = \frac{1}{M}\sum_{k=1}^{M} NB(t, \mathbf{X}_k) \tag{7}$$

(with $\mathbf{X}_k$ sampled from the distribution of $\mathbf{X}|X_i$), the estimator $U_t(X_i)$ is an unbiased estimator of $\mu_t(X_i)$, that is,

$$E_{\mathbf{X}|x_i}\{U_t(X_i)\} = \mu_t(X_i) \tag{8}$$

However,

$$E_{\mathbf{X}|x_i}\{\hat{m}(X_i)\} = E_{\mathbf{X}|x_i}\left[\max_t\{U_1(X_i), \ldots, U_T(X_i)\}\right] \tag{9}$$

$$\geq \max\{\mu_1(X_i), \ldots, \mu_T(X_i)\} \tag{10}$$

$$= m(X_i) \tag{11}$$

and so we would expect $\hat{m}(X_{i,j})$ to overestimate $m(X_{i,j})$ for any value of X_i. Equality only holds in equation (10) if the variances of the $U_t(X_i)$ variables are sufficiently small, that is, for some s we have $P(U_s(X_i) > U_t(X_i)) = 1$ for $t \neq s$. This can (effectively) be achieved if M is sufficiently large. Hence **both** N and M must be suitably large to ensure convergence to the true value of the partial EVPI.

The size of the bias will depend on the overlap of the sampling distributions of the $U_t(X_i)$s. For sufficiently large M we will have approximate normality:

$$U_t(X_i) \sim N\left[\mu_t(X_i), \frac{1}{\mathrm{M}}\tau^2(X_i)\right] \tag{12}$$

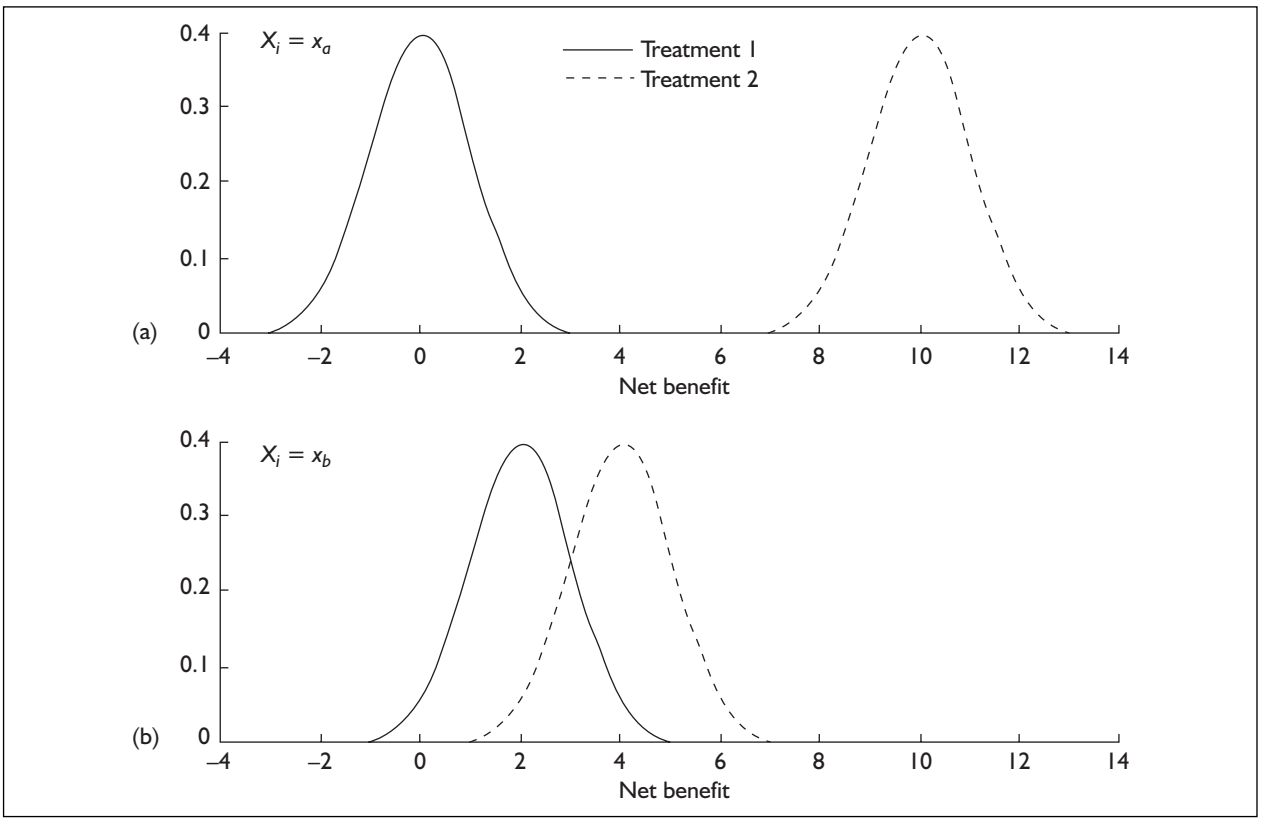

FIGURE 7 *In (a), the sampling distributions of* $U_1(X_i)$ *and* $U_2(X_i)$ *are sufficiently distinct for* $\hat{m}(X_i)$ *to be an (approximately) unbiased estimator of* $m(X_i)$. *In (b), the overlap means that on average* $\hat{m}(X_i)$ *will overestimate* $m(X_i)$.

with $\tau^2(X_i) = \text{Var}_{\mathbf{X}|x_i}\{\text{NB}(t, \mathbf{X})|X_i$. Now consider the diagrams in *Figure 7*.

Suppose we just have two treatment options. In *Figure 7(a)*, we suppose that conditional on $X_i = x_a$, the sampling distributions of $U_1(X_i)$ and $U_2(X_i)$ are the plotted functions. There is effectively no overlap between these distributions, and so we will have

$$E_{\mathbf{X}|x_i} \max\{U_1(X_i), U_2(X_i) \simeq \max\{\mu_1(X_i),\mu_2(X_i)\} \tag{13}$$

and the bias in the estimator $\hat{m}(X_i)$ will be negligible at $X_i = x_a$. Now consider *Figure 7(b)*. Here, at an alternative value x_b of X_i, we have substantial overlap between the sampling distributions of $U_1(X_i)$ and $U_2(X_i)$, and we will have

$$E_{\mathbf{X}|x_i}[\max\{U_1(X_i), U_2(X_i)\}| > \max\{\mu_1(X_i), \mu_2(X_i)\} \tag{14}$$

and the bias in the estimator $\hat{m}(X_i)$ will be more substantial at $X_i = x_b$.

As we increase M, the variance of each $U_t(X_i)$ will decrease, decreasing the size of the overlap when it does occur. However, if the partial EVPI of X_i is non-zero, there will always exist values of X_i which produce this overlap.

We now need to establish the minimum sample size M such that the bias in the estimator will be sufficiently small. If we choose a sufficiently large value of N for the first stage of the Monte Carlo sampling, then we just need to consider the expected bias, which is given by

$$E_{x_i}\left[E_{\mathbf{X}|x_i}\{\max_t U_t(X_i)\} - \max_t\{\mu_t(X_i)\}\right] \tag{15}$$

In principle, if we knew $\mu_t(X_i)$, $\tau^2(X_i)$ and $\text{Cov}\{U_s(X_i), U_t(X_i)\}$ for all X_i, we could then estimate equation (15) for any M using Monte Carlo, provided that M is sufficiently large for the normal approximation in equation (12) to hold:

1. Generate a random value of X_i.
2. Generate a random vector $\{U_1(X_i), \ldots U_T(X_i)\}$ from its joint (multivariate normal) distribution. Note that the variance–covariance matrix will be a function of M.
3. Calculate the sampled bias:

$$\max_t\{U_t(X_i)\} - \max_t\{(\mu_t(X_i)\} \tag{16}$$

4. Repeat steps 1–3 many times and compute the mean sampled bias.

By considering different sample sizes M, we can then find the smallest M that gives an acceptably small mean bias. The obvious difficulty with this is that to evaluate $\mu_t(X_i)$, $\tau^2(X_i)$ and $\text{Cov}\{U_s(X_i), U_t(X_i)\}$ accurately for any X_i, we would have to run the economic model many times, enough to obtain a good estimate of the partial EVPI in the first place. Consequently, the following approximate procedure is suggested:

1. Choose a small number of evenly spaced design points (say 21) $\mathbf{X}_{i1}, \ldots, \mathbf{X}_{i21}$ to cover the sample space $\mathbf{X}_i$.
2. For each X_{ij} generate a random sample $X_{ij1}, \ldots, X_{ij30}$ from the distribution of $X|X_{ij}$.
3. Evaluate $NB(t, \mathbf{X}_{ij1}), \ldots, NB(t, \mathbf{X}_{ij30})$ for each t and use the sample means and variance–covariance matrix to estimate $E\{\mathbf{U}(X_{ij})\}$ and $\text{Var}\{\mathbf{U}(X_{ij})\}$ [with $\mathbf{U} = (U_i, \ldots, U_T)^{\text{T}}$]. Denote these estimates by $\hat{\boldsymbol{\mu}}(X_{ij})$ and $\hat{V}(X_{ij})$.

The sample size of 21 is suggested as the minimum number needed for reliable numerical integration in the second stage given below. The sample size of 30 is designed to approximate estimates of the means and variances that will be of the right orders of magnitude. A larger number can of course be used, but the intention at this stage is to keep the number of model runs to a minimum.

We can now obtain an approximate estimate of the expected bias for any sample size M used in the second stage of the Monte Carlo calculation:

1. For a given M, approximate the distribution of $\mathbf{U}(X_{ij})$ by

$$\mathbf{U}(X_{ij}) \sim N\left[\hat{\mu}(X_{ij}), \frac{1}{M}\hat{V}(X_{ij})\right] \tag{17}$$

2. Generate a large a sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_K$ from this distribution and estimate the bias of $\hat{m}(X_{ij})$ by

$$\hat{b}(X_{ij}) = \frac{1}{K}\sum_{k=1}^{K}[\max(\mathbf{Z}_k) - \max\{\hat{\mu}(X_{ij})\}] \tag{18}$$

3. Repeat steps 1 and 2 for all $X_{i1}, \ldots, X_{i21}$.
4. Use numerical integration to estimate

$$E_{x_i}\left[E_{\mathbf{X}|x_i}\{\max_t U_t(X_i)\} - \max_t\{\mu t(X_i)\}\right] \simeq E_{x_i}\{\hat{b}(X_i)\} \tag{19}$$

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley,
Director, NHS HTA Programme, Department of Pharmacology & Therapeutics, University of Liverpool

Professor Bruce Campbell, Consultant Vascular & General Surgeon, Royal Devon & Exeter Hospital

Professor Shah Ebrahim, Professor in Epidemiology of Ageing, University of Bristol

Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Radcliffe Hospital, Oxford

Dr Ron Zimmern, Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley,
Director, NHS HTA Programme, Department of Pharmacology & Therapeutics, University of Liverpool

Chair,
Professor Shah Ebrahim,
Professor in Epidemiology of Ageing, Department of Social Medicine, University of Bristol

Deputy Chair,
Professor Jenny Hewison,
Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine

Dr Jeffrey Aronson
Reader in Clinical Pharmacology, Department of Clinical Pharmacology, Radcliffe Infirmary, Oxford

Professor Ann Bowling, Professor of Health Services Research, Primary Care and Population Studies, University College London

Professor Andrew Bradbury, Professor of Vascular Surgery, Department of Vascular Surgery, Birmingham Heartlands Hospital

Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health & Related Research, University of Sheffield

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York

Dr Andrew Farmer, Senior Lecturer in General Practice, Department of Primary Health Care, University of Oxford

Professor Fiona J Gilbert, Professor of Radiology, Department of Radiology, University of Aberdeen

Professor Adrian Grant, Director, Health Services Research Unit, University of Aberdeen

Professor F D Richard Hobbs, Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry, University of Cambridge

Professor Sallie Lamb, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health, Coventry University

Professor Julian Little, Professor of Epidemiology, Department of Medicine and Therapeutics, University of Aberdeen

Professor Stuart Logan, Director of Health & Social Care Research, The Peninsula Medical School, Universities of Exeter & Plymouth

Professor Tim Peters, Professor of Primary Care Health Services Research, Division of Primary Health Care, University of Bristol

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh

Professor Mark Sculpher, Professor of Health Economics, Centre for Health Economics, Institute for Research in the Social Services, University of York

Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Birmingham

Ms Kate Thomas, Deputy Director, Medical Care Research Unit, University of Sheffield

Professor Simon G Thompson, Director, MRC Biostatistics Unit, Institute of Public Health, Cambridge

Ms Sue Ziebland, Senior Research Fellow, Cancer Research UK, University of Oxford

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Ms Norma Armston, Freelance Consumer Advocate, Bolton

Professor Max Bachmann Professor Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia

Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust

Dr Paul Cockcroft, Consultant Medical Microbiologist/Laboratory Director, Public Health Laboratory, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, Addenbrooke's Hospital, Cambridge

Dr David Elliman, Consultant in Community Child Health, London

Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea

Dr John Fielding, Consultant Radiologist, Radiology Department, Royal Shrewsbury Hospital

Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen

Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Edmund Jessop, Medical Adviser, National Specialist Commissioning Advisory Group (NSCAG), Department of Health, London

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London

Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton

Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust

Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Head of Division of Reproductive & Child Health, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Oxford Radcliffe Hospital

Professor Tony Avery, Professor of Primary Health Care, University of Nottingham

Professor Stirling Bryan, Professor of Health Economics, Health Services Management Centre, University of Birmingham

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Dr Christopher Cates, GP and Cochrane Editor, Bushey Health Centre

Professor Imti Choonara, Professor in Child Health, University of Nottingham, Derbyshire Children's Hospital

Mr Charles Dobson, Special Projects Adviser, Department of Health

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff

Mrs Sharon Hart, Managing Editor, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, Bristol South & West Primary Care Trust

Professor Stan Kaye, Professor of Medical Oncology, Consultant in Medical Oncology/Drug Development, The Royal Marsden Hospital

Ms Barbara Meredith, Project Manager Clinical Guidelines, Patient Involvement Unit, NICE

Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London

Professor Jan Scott, Professor of Psychological Treatments, Institute of Psychiatry, University of London

Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry

Dr Helen Williams, Consultant Microbiologist, Norfolk & Norwich University Hospital NHS Trust

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and General Surgeon, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Head of Clinical Support & Health Protection, Directorate of Health and Social Care (North), Department of Health, Manchester

Dr Aileen Clarke,
Reader in Health Services Research, Public Health & Policy Research Unit,
Barts & the London School of Medicine & Dentistry,
Institute of Community Health Sciences, Queen Mary,
University of London

Mr Matthew William Cooke,
Senior Clinical Lecturer and Honorary Consultant,
Emergency Department,
University of Warwick, Coventry & Warwickshire NHS Trust,
Division of Health in the Community, Centre for Primary Health Care Studies, Coventry

Dr Carl E Counsell, Senior Lecturer in Neurology,
University of Aberdeen

Dr Keith Dodd, Consultant Paediatrician, Derbyshire Children's Hospital

Professor Gene Feder, Professor of Primary Care R&D, Barts & the London, Queen Mary's School of Medicine and Dentistry, University of London

Professor Paul Gregg,
Professor of Orthopaedic Surgical Science, Department of Orthopaedic Surgery,
South Tees Hospital NHS Trust

Ms Bec Hanley, Freelance Consumer Advocate,
Hurstpierpoint

Ms Maryann L. Hardy,
Lecturer,
Division of Radiography,
University of Bradford

Professor Alan Horwich,
Director of Clinical R&D, The Institute of Cancer Research,
London

Dr Phillip Leech, Principal Medical Officer for Primary Care, Department of Health,
London

Dr Simon de Lusignan,
Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences,
St George's Hospital Medical School, London

Dr Mike McGovern, Senior Medical Officer, Heart Team,
Department of Health, London

Professor James Neilson,
Professor of Obstetrics and Gynaecology, Dept of Obstetrics and Gynaecology,
University of Liverpool,
Liverpool Women's Hospital

Dr John C Pounsford,
Consultant Physician, North Bristol NHS Trust

Dr Vimal Sharma,
Consultant Psychiatrist & Hon Snr Lecturer,
Mental Health Resource Centre,
Victoria Central Hospital,
Wirrall

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh,
Professor of Public Health,
University of Aberdeen

Expert Advisory Network

Members

Professor Douglas Altman, Director of CSM & Cancer Research UK Med Stat Gp, Centre for Statistics in Medicine, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor John Bond, Director, Centre for Health Services Research, University of Newcastle upon Tyne, School of Population & Health Sciences, Newcastle upon Tyne

Mr Shaun Brogan, Chief Executive, Ridgeway Primary Care Group, Aylesbury

Mrs Stella Burnside OBE, Chief Executive, Office of the Chief Executive. Trust Headquarters, Altnagelvin Hospitals Health & Social Services Trust, Altnagelvin Area Hospital, Londonderry

Ms Tracy Bury, Project Manager, World Confederation for Physical Therapy, London

Mr John A Cairns, Professor of Health Economics, Health Economics Research Unit, University of Aberdeen

Professor Iain T Cameron, Professor of Obstetrics and Gynaecology and Head of the School of Medicine, University of Southampton

Dr Christine Clark, Medical Writer & Consultant Pharmacist, Rossendale

Professor Collette Mary Clifford, Professor of Nursing & Head of Research, School of Health Sciences, University of Birmingham, Edgbaston, Birmingham

Professor Barry Cookson, Director, Laboratory of Healthcare Associated Infection, Health Protection Agency, London

Professor Howard Stephen Cuckle, Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, University of York

Dr Katherine Darton, Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux, Professor of Paediatric Epidemiology, London

Mr John Dunning, Consultant Cardiothoracic Surgeon, Cardiothoracic Surgical Unit, Papworth Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw, Consultant Vascular Surgeon, Gloucestershire Royal Hospital, Gloucester

Professor Martin Eccles, Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby, Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick, Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Professor David Field, Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust

Mrs Gillian Fletcher, Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield

Professor Jayne Franklyn, Professor of Medicine, Department of Medicine, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham

Ms Grace Gibbs, Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Servs, West Middlesex University Hospital, Isleworth

Dr Neville Goodman, Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Alastair Gray, Professor of Health Economics, Department of Public Health, University of Oxford

Professor Robert E Hawkins, CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor F D Richard Hobbs, Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Allen Hutchinson, Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Dr Duncan Keeley, General Practitioner (Dr Burch & Ptnrs), The Health Centre, Thame

Dr Donna Lamping, Research Degrees Programme Director & Reader in Psychology, Health Services Research Unit, London School of Hygiene and Tropical Medicine, London

Mr George Levvy, Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay, Professor of Psychiatry for the Elderly, University of Leicester, Leicester General Hospital

Professor Rajan Madhok, Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor David Mant, Professor of General Practice, Department of Primary Care, University of Oxford

Professor Alexander Markham, Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Dr Chris McCall, General Practitioner, The Hadleigh Practice, Castle Mullen

Professor Alistair McGuire, Professor of Health Economics, London School of Economics

Dr Peter Moore, Freelance Science Writer, Ashtead

Dr Andrew Mortimore, Consultant in Public Health Medicine, Southampton City Primary Care Trust

Dr Sue Moss, Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton

Professor Jon Nicholl, Director of Medical Care Research Unit, School of Health and Related Research, University of Sheffield

Mrs Julietta Patnick, National Co-ordinator, NHS Cancer Screening Programmes, Sheffield

Professor Robert Peveler, Professor of Liaison Psychiatry, University Mental Health Group, Royal South Hants Hospital, Southampton

Professor Chris Price, Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Ms Marianne Rigge, Director, College of Health, London

Dr Eamonn Sheridan, Consultant in Clinical Genetics, Genetics Department, St James's University Hospital, Leeds

Dr Ken Stein, Senior Clinical Lecturer in Public Health, Director, Peninsula Technology Assessment Group, University of Exeter

Professor Sarah Stewart-Brown, Director HSRU/Honorary Consultant in PH Medicine, Department of Public Health, University of Oxford

Professor Ala Szczepura, Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor, Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster, Consumer member, HTA – Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278