# A017: Fast efficient computation of Value of Information from a Probabilistic Sensitivity Analysis sample: a non-parametric regression approach

Mark Strong[1§], Jeremy E. Oakley[2], Alan Brennan[1], Penny Watson[1]

1. School of Health and Related Research (ScHARR)

2. School of Mathematics and Statistics

University of Sheffield, 30 Regent Street, Sheffield S1 4DA, UK.

[§] Corresponding author: email m.strong@sheffield.ac.uk; tel +44 (0)114 222 0812.

**Abstract**

**Background:** Health economic models are used to estimate the expected net benefits of competing decision options. The true values of the input parameters of such models are rarely known with certainty, and it is often useful to quantify the value of undertaking further data collection in order to reduce uncertainty. An upper bound on the value of learning a subset of input parameters is quantified by its partial Expected Value of Perfect Information (EVPI). The value of a particular data collection exercise is quantified by its Expected Value of Sample Information (EVSI). Delaying an adoption decision in order to collect new data is optimal if the cost of data collection is less than the expected value of that data collection exercise.

**Problems when computing partial EVPI and EVSI:** The standard approach to computing both partial EVPI and EVSI is via a nested two-level Monte Carlo scheme that includes at each inner loop step both parameter sampling and model evaluation. This scheme can be prohibitively slow for complex models, particularly those that require for each model run a large number of patient-level simulation steps. Additional problems arise if the two-level Monte Carlo scheme results in an inner loop conditional distribution that is difficult to sample from. This most commonly occurs when computing EVSI for a problem in which the parameter distribution is not conjugate to the data likelihood, but can also occur when computing partial EVPI where parameters are correlated. In either case we typically need to resort to Markov Chain Monte Carlo methods, implemented for example in WinBUGS. In practice, these difficulties have resulted in the restriction of Value of Information analyses to only a small subset of health economic evaluation studies.

**Proposed solution:** To overcome the problems above we present novel, fast and efficient non-parametric regression based methods for computing partial EVPI and EVSI. The methods require only the 'probabilistic sensitivity analysis' (PSA) sample: a single set of samples from the model parameters, along with the corresponding model evaluations. The new methods allow Value of Information measures to be computed for models of any complexity, and hence be made more widely available to modellers and decision makers.

## Acknowledgements

Mark Strong is supported by a post-doctoral NIHR fellowship. This report is independent research supported by the National Institute for Health Research (Post-Doctoral Fellowship, PDF-2012-05-258). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

# 1 Introduction

***What is Value of Information (VoI) in the context of health economic decision making? What is partial EVPI and what is EVSI? How do we calculate VoI? What are the problems with the Monte Carlo approach to calculating VoI?***

## 1.1 Health Economic Decision Modelling and Value of Information (VoI)

Health economic decision analytic models are used to estimate the expected net benefits of competing decision options. The true values of the input parameters of such models are rarely known with certainty, and it is often useful to quantify the value to the decision maker of reducing uncertainty about the model parameters. The value of learning, with certainty, a parameter (or a group of parameters) can be quantified by its partial expected value of perfect information (partial EVPI) (Raiffa, 1968). The partial EVPI for a parameter reveals the sensitivity of the decision to our uncertainty about that parameter.

We will, however, rarely be able to learn any parameter with certainty, and EVPI will typically therefore only provide an upper bound on the value of information for any particular parameter. What we really want to know is the value of conducting further research before making the resource allocation decision in question. The value of *reducing*, rather than *eliminating*, uncertainty through the collection of data is captured by the expected value of sample information (EVSI), and it is therefore this measure that provides a realistic assessment of the value of future research, not the EVPI.

The use of health economic decision modelling to inform adoption and reimbursement decisions is well established in the process of appraisal used by the National Institute for Health and Care Excellence (NICE). However, the experience of using similarly explicit methods to inform research decisions is less well developed. Uncertainty in the reimbursement decision may motivate decision makers to consider options for further data collection alongside the adoption of new technologies, or to delay adoption until after data collection. It has been shown that delaying a decision until further data is collect can be optimal if the opportunity costs are potentially large (McKenna and Claxton, 2011). The application of EVSI analysis to evaluate alternative research designs in this reimbursement setting has been limited.

## 1.2 Computation of Partial EVPI

The partial EVPI for a single parameter (or group of parameters) of interest is typically calculated via a two-level nested Monte Carlo approach. This requires us to sample values of the input parameter(s) of interest in an outer loop, and then to sample values from the joint conditional distribution of the remaining parameters and run the model in an inner loop (Brennan et al., 2007). We recognise three important limitations to this method. Firstly, the two-level method is computationally demanding for all but very simple models due to the nested loop scheme. Secondly, the approach requires that the model is run as part of the EVPI calculation process, which may be difficult if the VoI analysis is to be conducted by someone who does not have access to the decision model. Lastly, a potential problem arises in cases where correlations exist between parameters. If the parameters of interest are correlated with the remaining parameters then for the two-level Monte Carlo method to work, there must be some method of

sampling from the distribution of the remaining parameters, *conditional* on the values of the parameters of interest that have been sampled in the outer loop. If the required conditional distributions are difficult to sample from, say requiring Markov chain Monte Carlo (MCMC), then the computational burden will be substantially further increased.

Our experience is that, whilst probabilistic sensitivity analysis (PSA) has become the norm in many economic evaluations for health technology assessment, it is much less common for partial EVPIs to be estimated. In our view the reasons for this are partly technical (in terms of the extra demands on the statistical and programming skills of the analyst), partly computational (the additional model development and model running time to implement two nested loops re-running the model on each iteration), and partly structural (in that decision makers and research funding bodies have not always demanded these analyses).

Recently, computationally efficient methods for calculating partial EVPI for single parameters have been published, (Strong et al., 2012; Sadatsafavi et al., 2013). The restriction to single parameters is potentially problematic since we often expect research to update our knowledge about groups of parameters (for example a set of relative risks, or a group of related costs) rather than just single parameters. To address this limitation we have recently published a much more general approach to computing partial EVPI for any number of parameters (Strong et al., 2013b). It is this work, and the extension of this work to EVSI, that we present in here.

### 1.3 Computation of EVSI

The general approach to computing EVSI is very similar to that of computing partial EVPI. A nested two-level Monte Carlo scheme is employed in which plausible datasets are generated in an outer loop, and then conditional on each dataset, samples are generated from the posterior distribution of the parameters in an inner loop. Again, the model is run at each iteration of the inner loop, and thus this scheme can easily become computationally burdensome. The two-level Monte Carlo approach will also be difficult if the prior distribution of the model parameters is not conjugate to the data likelihood. In this case generating the inner loop samples will typically require Markov chain Monte Carlo (MCMC), and the repeated application of MCMC for each sampled dataset adds considerably to the computational burden. Computationally simpler approaches are sometimes available, but these rely either on the model being of a certain form (Ades et al., 2004; Welton et al., 2013), or on assumptions of Normality of the mean incremental net benefits (Eckermann and Willan, 2007).

The concept of EVSI was first discussed in the health economics literature well over a decade ago (key papers are Claxton and Posnett, 1996; Claxton, 1999a,b). Despite this, very few economic evaluation studies report EVSI. This reflects, at least in part, the computational burden of the two level Monte Carlo scheme. For example, in a recently published cost-effectiveness study, the authors noted that in order to compute EVSI without assuming an approximation that the model was linear, it would have taken 7.5 days (Soares et al., 2012). In another example the proposed EVSI analysis would have taken 37.5 days (Brennan et al., 2007). Clearly, computation times of this order are prohibitive.

### 1.4 Our proposed method

In this paper we propose a non-parametric regression based method that can be used to calculate both the partial EVPI of any subset of model parameters, and the EVSI of any data collection exercise. The method overcomes the important limitations of the two-level Monte Carlo methods described above. Our method requires only the single set of model evaluations that is generated in a standard probability sensitivity analysis (PSA). It does not require access to the decision model itself. The method makes no assumptions regarding the form of the model, makes no assumptions regarding the form of the joint distribution of the parameters, does not require the use of MCMC and in the case of EVSI, does not require that the parameter prior is conjugate to the data likelihood.

The paper is structured as follows. In Section 2 we define the decision problem, the model and measures of VoI. In section 3 we describe the typical Monte Carlo approach to computing VoI and discuss the limitations of this approach. In section 4 we introduce the non-parametric regression method. In section 5 we illustrate the method in two case studies. Both case study models have been used before for illustrative purposes (Ades et al., 2004; Brennan et al., 2007; Oakley et al., 2010; Kharroubi et al., 2011; Strong et al., 2012). In Section 6, we conclude with a discussion of the implications and limitations of the approach.

## 2 Definitions and notation

***How is partial EVPI defined? How is EVSI defined***

### 2.1 Notation for the decision problem and our model

We assume that we are faced with $D$ decision options, indexed $d = 1, \ldots, D$. Costs associated with option $d$ are $C_d$ and health outcomes are $E_d$. The monetary net benefit of option $d$ is given by $NB_d = \lambda E_d - C_d$ where $\lambda$ is the value to the decision maker of one unit of health outcome. Costs and health effects, and therefore net benefits, are unknown and we have therefore built a model $\mathrm{NB}(d, \boldsymbol{\theta})$ that aims to predict the net benefit of decision option $d$ given a vector of $r$ input parameter values $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r, \lambda)$. The true values of some or all of the parameters are assumed to be unknown and we represent beliefs about the input parameters via their joint distribution $p(\boldsymbol{\theta})$.

### 2.2 The overall expected value of perfect information

The overall expected value of information is the expected difference between the value of the optimal decision based on perfect information about all input parameters of the model, and the value of the decision made only with current information (Felli and Hazen, 1998, 2003).

The expected value of our optimal decision, made only with current information is

$$\max_d \mathbb{E}_{\boldsymbol{\theta}}\{\mathrm{NB}(d, \boldsymbol{\theta})\}. \tag{1}$$

If we knew the value of the model parameters then the net benefit of the optimal decision would be $\max_d \mathrm{NB}(d, \boldsymbol{\theta})$, but since $\boldsymbol{\theta}$ is unknown, we must average over our current information about

$\boldsymbol{\theta}$, giving

$$\mathbb{E}_{\boldsymbol{\theta}} \left\{ \max_d \text{NB}(d, \boldsymbol{\theta}) \right\}. \tag{2}$$

The overall EVPI is the difference between equation (2) and equation (1),

$$\text{EVPI}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}} \left\{ \max_d \text{NB}(d, \boldsymbol{\theta}) \right\} - \max_d \mathbb{E}_{\boldsymbol{\theta}} \{\text{NB}(d, \boldsymbol{\theta})\}. \tag{3}$$

### 2.3 The partial expected value of perfect information

The partial expected value of information is the expected difference between the value of the optimal decision based on perfect information about some subset of inputs and the value of the decision made only with current information (Felli and Hazen, 1998, 2003).

We denote the vector of $p$ input parameters for which we wish to calculate the partial EVPI as $\boldsymbol{\theta}_i$ and the remaining $r - p$ input parameters as $\boldsymbol{\theta}_{-i}$. We denote the expectation over the marginal distribution of $\boldsymbol{\theta}_i$ as $\mathbb{E}_{\boldsymbol{\theta}_i}$, and over the conditional distribution of $\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i$ as $\mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i}$.

If we knew the value of the inputs of interest, $\boldsymbol{\theta}_i$, then the optimal decision would be that with the greatest net benefit, after averaging over the conditional distribution of the remaining unknown inputs $\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i$. The expected net benefit would be

$$\max_d \mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i} \{\text{NB}(d, \boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i})\}. \tag{4}$$

But, since $\boldsymbol{\theta}_i$ is unknown, we must average over our current information about $\boldsymbol{\theta}_i$ giving

$$\mathbb{E}_{\boldsymbol{\theta}_i} \left[ \max_d \mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i} \{\text{NB}(d, \boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i})\} \right]. \tag{5}$$

The partial EVPI for inputs $\boldsymbol{\theta}_i$ is the difference between equation (5) and equation (1),

$$\text{EVPI}(\boldsymbol{\theta}_i) = \mathbb{E}_{\boldsymbol{\theta}_i} \left[ \max_d \mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i} \{\text{NB}(d, \boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i})\} \right] - \max_d \mathbb{E}_{\boldsymbol{\theta}} \{\text{NB}(d, \boldsymbol{\theta})\}. \tag{6}$$

We are commonly in a situation in which we cannot evaluate any of the three expectations in equation (6) analytically. Important exceptions are cases in which models are either of linear form (e.g., NB $= \beta_1\theta_1 + \beta_2\theta_2$) or multilinear (sum-product) form (e.g., NB $= \beta_1\theta_1\theta_2 + \beta_2\theta_3\theta_4$) (where $\beta_1$ and $\beta_2$ are constants). In the linear case, the expectation in the second term and the inner expectation in the first term both have an analytic solution, and in the multilinear case, these expectations have an analytic solution if inputs are independent (Ades et al., 2004). In the case of correlated inputs, analytic solutions to these two expectations will sometimes exist, such as the case in which the inputs have a multivariate Normal distribution. The outer expectation in the first term is more problematic due to the maximization step, and analytic solutions rarely exist. See Brennan et al. (2007) for a fuller discussion.

At this point we note that we can re-express equation (6) as

$$\text{EVPI}(\boldsymbol{\theta}_i) = \mathbb{E}_{\boldsymbol{\theta}_i} \left[ \max_d \mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i} \{\text{NB}(d, \boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i})\} \right] - \max_d \mathbb{E}_{\boldsymbol{\theta}_i} \left[ \mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i} \{\text{NB}(d, \boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i})\} \right]. \tag{7}$$

The reason for the re-expression will become apparent when we discuss Monte Carlo sampling schemes for estimating partial EVPI.

### 2.4 The expected value of sample information

The expected value of sample information (EVSI) is the expected difference between the value of the optimal decision based on some sample of data, informative for some subset of inputs, and the value of the decision made only with prior information (Raiffa, 1968).

We envisage that we can collect data that will be informative for some subset of parameters, $\boldsymbol{\theta}_i$ where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i}\}$. We consider the (as yet uncollected) data as a vector of random variables, and denote this as upper case $\mathbf{X}$. We note that if $\mathbf{X}$ is only informative for $\boldsymbol{\theta}_i$ then we have in effect stated that $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{-i}$ are *a priori* independent.

The expected value of our optimal decision, made only with current information is given by equation (1). If we had data $\mathbf{X}$ that were informative for $\boldsymbol{\theta}_i$, then the optimal decision would be that with the greatest net benefit, after averaging over the joint distribution of the inputs conditional on the data, $p(\boldsymbol{\theta}|\mathbf{X})$. The expected net benefit would be

$$\max_d \mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}} \left\{ \mathrm{NB}(d, \boldsymbol{\theta}) \right\}. \tag{8}$$

We note that, because $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{-i}$ are independent, $p(\boldsymbol{\theta}|\mathbf{X})$ factorises as $p(\boldsymbol{\theta}_i|\mathbf{X})p(\boldsymbol{\theta}_{-i})$. The data $\mathbf{X}$ are, however, uncollected and we must therefore average over possible datasets,

$$\mathbb{E}_{\mathbf{X}} \left[ \max_d \mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}} \left\{ \mathrm{NB}(d, \boldsymbol{\theta}) \right\} \right]. \tag{9}$$

The EVSI for data collection exercise $\mathbf{X}$ is then the difference between equation (9) and equation (1),

$$\mathrm{EVSI}(\mathbf{X}) = \mathbb{E}_{\mathbf{X}} \left[ \max_d \mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}} \left\{ \mathrm{NB}(d, \boldsymbol{\theta}) \right\} \right] - \max_d \mathbb{E}_{\boldsymbol{\theta}} \{ \mathrm{NB}(d, \boldsymbol{\theta}) \}. \tag{10}$$

As with the re-expression of equation (6) as equation (7), we can re-express equation (10) as

$$\mathrm{EVSI}(\mathbf{X}) = \mathbb{E}_{\mathbf{X}} \left[ \max_d \mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}} \left\{ \mathrm{NB}(d, \boldsymbol{\theta}) \right\} \right] - \max_d \mathbb{E}_{\mathbf{X}} \left[ \mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}} \{ \mathrm{NB}(d, \boldsymbol{\theta}) \} \right]. \tag{11}$$

The reason for the re-expression will become apparent when we discuss Monte Carlo schemes.

## 3 Monte Carlo approaches to computing VoI

***What is probability sensitivity analysis (PSA)? How do we compute overall and partial EVPI? How do we compute EVSI? What are the problems with the typical Monte Carlo approach to computing VoI?***

### 3.1 Probability sensitivity analysis

We denote the $n^{th}$ sample drawn from the joint distribution of the input parameters as $\boldsymbol{\theta}^{(n)} = (\theta_1^{(n)}, \ldots, \theta_r^{(n)})$. Note that we use bracketed superscript notation to index the sample set.

A 'probabilistic sensitivity analysis' (PSA) takes $N$ samples from the joint distribution of the input parameters, $\{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}\}$, and generates a corresponding set of $N$ net benefits $\{\mathrm{NB}(d, \boldsymbol{\theta}^{(1)}), \ldots, \mathrm{NB}(d, \boldsymbol{\theta}^{(N)})\}$ for each decision option $d$. The PSA allows us estimate the expected value of each decision option via Monte Carlo, and hence find the value of the optimal option,

$$\max_d \mathbb{E}_\theta \{ \mathrm{NB}(d, \boldsymbol{\theta}) \} \simeq \max_d \frac{1}{N} \sum_{n=1}^{N} \mathrm{NB}(d, \boldsymbol{\theta}^{(n)}). \tag{12}$$

The PSA also allows us to the estimate the overall EVPI, i.e. the value of learning *all* the parameters of the model, via

$$\widehat{\text{EVPI}}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\max_{d}\text{NB}(d,\boldsymbol{\theta}^{(n)}) - \max_{d}\frac{1}{N}\sum_{n=1}^{N}\text{NB}(d,\boldsymbol{\theta}^{(n)}). \tag{13}$$

### 3.2 The Monte Carlo approach to computing partial EVPI

The Monte Carlo method for estimating partial EVPI involves a nested two-level approach with $K$ 'outer' simulations and $J$ 'inner' simulations. The estimator is given by

$$\frac{1}{K}\sum_{k=1}^{K}\max_{d}\frac{1}{J}\sum_{j=1}^{J}\text{NB}\left(d,\boldsymbol{\theta}_i^{(k)},\boldsymbol{\theta}_{-i}^{(j,k)}\right) - \max_{d}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{J}\sum_{j=1}^{J}\text{NB}\left(d,\boldsymbol{\theta}_i^{(k)},\boldsymbol{\theta}_{-i}^{(j,k)}\right), \tag{14}$$

where $\boldsymbol{\theta}_i^{(k)}$ are samples drawn from the prior distribution of the parameters of interest, $p(\boldsymbol{\theta}_i)$, and $\boldsymbol{\theta}_{-i}^{(j,k)}$ are samples drawn from the conditional distribution $p(\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i^{(k)})$.

Note that our estimator follows the structure of equation (7) rather than the structure of equation (6). This allows us to exploit the positive correlation between the two terms in equation (14) and hence produce an estimator with lower variance (Brennan et al., 2007).

### 3.3 The Monte Carlo approach to computing EVSI

The Monte Carlo approach to estimating EVSI is similar to that for estimating partial EVPI. The important difference is that we need some means to simulate datasets $\mathbf{X}$ that may arise given the study that we propose. First, we must postulate a data generating process, i.e. we must propose a probability distribution for the data, conditional on the model parameters, $p(\mathbf{X}|\boldsymbol{\theta})$. Once we have this, the distribution of $\mathbf{X}$ can be obtained by the marginalisation of $p(\mathbf{X},\boldsymbol{\theta}) = p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, and this expression suggests a straightforward Monte Carlo sampling scheme for $\mathbf{X}$. We sample first a value $\boldsymbol{\theta}^*$ from the prior $p(\boldsymbol{\theta})$ and then sample $\mathbf{X}$ from the data density $p(\mathbf{X}|\boldsymbol{\theta}=\boldsymbol{\theta}^*)$. We recognise that, because $\mathbf{X}$ is only informative for $\boldsymbol{\theta}_i$, and therefore that $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{-i}$ are independent, that this is equivalent to sampling $\boldsymbol{\theta}_i^*$ from the prior $p(\boldsymbol{\theta}_i)$ and then sampling $\mathbf{X}$ from the data likelihood $p(\mathbf{X}|\boldsymbol{\theta}_i=\boldsymbol{\theta}_i^*)$.

The two-level Monte Carlo EVSI estimator is given by

$$\widehat{\text{EVSI}}(\mathbf{X}) = \frac{1}{K}\sum_{k=1}^{K}\max_{d}\frac{1}{J}\sum_{j=1}^{J}\text{NB}(d,\boldsymbol{\theta}^{(j,k)}) - \max_{d}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{J}\sum_{j=1}^{J}\text{NB}(d,\boldsymbol{\theta}^{(j,k)}), \tag{15}$$

where $\boldsymbol{\theta}^{(j,k)}$ are samples drawn from the posterior distribution of $\boldsymbol{\theta}|\mathbf{x}^{(k)}$, and $\mathbf{x}^{(k)}$ are generated by first sampling $\boldsymbol{\theta}^{(k)}$ from $p(\boldsymbol{\theta})$ and then $\mathbf{x}^{(k)}$ from $p(\mathbf{X}|\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)})$.

Note that our estimator follows the structure of equation (11) rather than (10). As with the partial EVPI estimator, this allows us to exploit the positive correlation between the two terms in equation (15) and hence produce an estimator with lower variance.

### 3.4 Problems with the 2-level Monte Carlo method

The first problem with the nested two-level scheme is the requirement to evaluate the net benefit function (i.e. to run the model) at each iteration of the inner loop, resulting in $J \times K$ model

evaluations. Sufficient numbers of runs of both the outer and inner loops are required to ensure that the partial EVPI is estimated with sufficient precision, and with an acceptable level of upward bias that is induced by the maximisation step (Oakley et al., 2010). For models that are slow to run this two-level scheme can represent a considerable computational burden. If the model is slow to run, and/or if $J$ and $K$ are large (in order to obtain adequate precision and minimal bias), then the scheme will be computationally burdensome.

A second potential problem is the requirement to sample from potentially difficult conditional distributions. In the case of partial EVPI, if the parameters of interest are correlated with the remaining parameters then there must be some method of sampling from the distribution of the remaining parameters, *conditional* on the values of the parameters of interest that have been sampled in the outer loop, i.e. from $p(\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i)$. In the case of EVSI we must be able to sample from $p(\boldsymbol{\theta}|\mathbf{X})$. If these conditional distributions require MCMC then this will add considerably to the computational burden (the MCMC step will occur $K$ times, each chain being of length $J$). Setting up the MCMC sampler (e.g. via writing WinBUGS code) and checking the MCMC chain(s) for convergence also requires investment in modeller time.

# 4 Estimating VoI using non-parameteric regression

***How can we frame our problem of estimating partial EVPI and EVSI as a non-parametric regression problem?***

## 4.1 Principles of estimating partial EVPI using regression

Firstly, we restate the partial EVPI expression,

$$\text{EVPI}(\boldsymbol{\theta}_i) = \mathbb{E}_{\boldsymbol{\theta}_i}\left[\max_d \mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i}\left\{\text{NB}(d,\boldsymbol{\theta}_i,\boldsymbol{\theta}_{-i})\right\}\right] - \max_d \mathbb{E}_{\boldsymbol{\theta}_i}\left[\mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i}\{\text{NB}(d,\boldsymbol{\theta}_i,\boldsymbol{\theta}_{-i})\}\right]. \quad (16)$$

In order to address the problems of the two-level method we focus our attention on the estimation of the inner expectation, $\mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i}\{\text{NB}(d,\boldsymbol{\theta}_i,\boldsymbol{\theta}_{-i})\}$. To avoid the need to compute this via Monte Carlo we re-frame the estimation of this expectation as a regression problem. We undertake three conceptual moves.

Firstly, we imagine that we have a PSA sample of size $N$, indexed $n = 1,\ldots,N$. We recognise that we can express the model output for model run $n$ as a sum of the conditional expectation that we require, and a mean-zero error term,

$$\text{NB}(d,\boldsymbol{\theta}^{(n)}) = \mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i^{(n)}}\{\text{NB}(d,\boldsymbol{\theta}_i^{(n)},\boldsymbol{\theta}_{-i})\} + \varepsilon^{(n)}. \quad (17)$$

The second move is to realise that the expectation $\mathbb{E}_{\boldsymbol{\theta}_{-i}|\boldsymbol{\theta}_i^{(n)}}\{\text{NB}(d,\boldsymbol{\theta}_i^{(n)},\boldsymbol{\theta}_{-i})\}$ takes a different value for each value $\boldsymbol{\theta}_i^{(n)}$ and can therefore be thought of as a function of $\boldsymbol{\theta}_i$. We do not know the form of this function, but we can denote it as the unknown function $g(d,\boldsymbol{\theta}_i)$, allowing us to write, for the $n^{th}$ model run,

$$\text{NB}(d,\boldsymbol{\theta}^{(n)}) = g(d,\boldsymbol{\theta}_i^{(n)}) + \varepsilon^{(n)}. \quad (18)$$

The third key idea is this. For each decision option $d \in \{1,\ldots,D\}$, we treat the net benefits $\{\text{NB}(d,\boldsymbol{\theta}^{(1)}),\ldots,\text{NB}(d,\boldsymbol{\theta}^{(N)})\}$ as 'noisy' data through which we can learn about the target

function $g(d, \boldsymbol{\theta}_i)$. Thus, we can think of this as $D$ regression problems. However, we immediately recognise that the target functions $g(d, \boldsymbol{\theta}_i)$ have unknown form, and we have no desire to impose any particular form. We could begin by fitting a standard linear model, with power and interaction terms to model the non-linearity between the net benefits and the inputs of interest, but we choose instead to adopt the more flexible 'non-parametric' regression approach offered by the Generalised Additive Model (GAM). GAM models assume that the expectation of the dependent variable is a smooth, but usually unknown, function of the independent variable, which is exactly what we need here. For an introduction to GAM models see Hastie and Tibshirani (1986) or Wood (2006). The models are easy to fit using the freely available software R (R Development Core Team, 2013).

We note here that partial EVPI is invariant to the re-expression of net benefits as incremental net benefits, relative to some chosen 'baseline' option. Under this re-expression the (incremental) net benefit of the baseline option is zero. This reduces the number of regression equations from $D$ to $D - 1$.

### 4.1.1 Hypothetical example

As an illustration, Figure 1 shows the results from a hypothetical PSA in which we plot the net benefit function, NB = NB$(\theta_1, \theta_2, \theta_3)$, against a single parameter of interest, $\theta_1$. The scatter of points suggests some kind of 'U-shaped' function. The dashed line shows a non-parametric regression of NB on $\theta_1$. This regression provides an estimate of the expected value $\mathbb{E}_{\theta_2,\theta_3|\theta_1}\text{NB}(\theta_1, \theta_2, \theta_3)$ as a function of $\theta_1$, i.e. it provides the $g(d, \boldsymbol{\theta}_i)$ from equation (18). In this particular illustrative model the expectation $\mathbb{E}_{\theta_2,\theta_3|\theta_1}\text{NB}(\theta_1, \theta_2, \theta_3)$ can be obtained analytically (solid line), showing that the true expectation is very well estimated by the non-parametric regression.

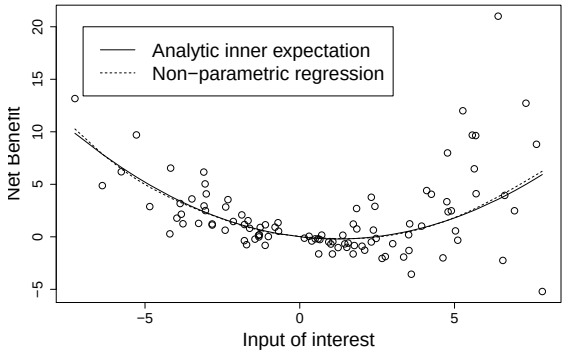


Figure 1: Net benefit against single input parameter of interest for hypothetical model with three parameters.

### 4.1.2 Partial EVPI calculation

After fitting a GAM model for each decision option $d$ we then extract the regression model fitted values. The fitted values are estimates of $g(d, \boldsymbol{\theta}_i^{(1)}), \ldots, g(d, \boldsymbol{\theta}_i^{(N)})$, our target quantity. We denote the GAM fitted values for decision option $d$ as $\hat{g}_d^{(1)}, \ldots, \hat{g}_d^{(N)}$, and the estimated

partial EVPI is then given by

$$\widehat{\text{EVPI}}(\boldsymbol{\theta}_i) = \frac{1}{N}\sum_{n=1}^{N}\max_d \hat{g}_d^{(n)} - \max_d \frac{1}{N}\sum_{n=1}^{N}\hat{g}_d^{(n)}. \tag{19}$$

The code in Box 1 illustrates the simplicity of the GAM regression approach using the `mgcv` package in R (R Development Core Team, 2013). In the example there are two decision options, with the vector object `INB` holding the incremental net benefits from the PSA. The PSA samples from the two parameters of interest are held in vector objects `theta1` and `theta2`. We assume the parameters do not interact non-additively in the model. If they did we would simply replace the model formula `INB ~ s(theta1)+s(theta2)` with the tensor product multivariate specification `INB ~ te(theta1,theta2)`.

**Box 1 - example R code for estimating partial EVPI via GAM regression**

```
library(mgcv)
model <- gam(INB ~ s(theta1) + s(theta2))
g.hat <- model$fitted
partial.evpi <- mean(pmax(0, g.hat)) - max(0, mean(g.hat))
```

### 4.2 Non-parametric regression method for EVSI

We restate the EVSI for data $\mathbf{X}$,

$$\text{EVSI}(\mathbf{X}) = \mathbb{E}_{\mathbf{X}}\left[\max_d \mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}}\left\{\text{NB}(d,\boldsymbol{\theta})\right\}\right] - \max_d \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}}\{\text{NB}(d,\boldsymbol{\theta})\}\right]. \tag{20}$$

The regression approach to computing EVSI follows closely that for calculating partial EVPI. Again, we focus our attention on the inner conditional expectation $\mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}}\{\text{NB}(d,\boldsymbol{\theta})\}$, and recognise that this can be thought of as an unknown function of $\mathbf{X}$. We denote this function $g(d,\mathbf{X})$, and write

$$\text{NB}(d,\boldsymbol{\theta}) \;=\; \mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}}\{\text{NB}(d,\boldsymbol{\theta})\} + \varepsilon \;=\; g(d,\mathbf{X}) + \varepsilon, \tag{21}$$

where $\varepsilon$ is a mean zero error term.

In some instances the data $\mathbf{X}$ will be high dimensional (e.g. censored time-to-event data in a study that measures survival), and if this is so we write the function $g$ in terms of some low dimensional summary statistic of the data $T(\mathbf{X})$ giving $\text{NB}(d,\boldsymbol{\theta}) = g\{d, T(\mathbf{X})\} + \varepsilon$. We discuss choice of summary statistic in the next section.

As before, for each decision option $d \in \{1,\ldots,D\}$, we treat the net benefits $\text{NB}(d,\boldsymbol{\theta})$ as 'noisy' data through which we can learn about the target function $g\{d,T(\mathbf{X})\}$. Thus, we can think of this as $D$ non-parametric regression problems. Obtaining the necessary 'data' for the regression analysis proceeds as follows. We assume we have at our disposal a PSA sample of size $N$. This consists of a set of samples, $\boldsymbol{\theta}^{(n)}$ $(n = 1,\ldots,N)$ from the distribution of the input parameters, and corresponding evaluations of the model $\text{NB}(d,\boldsymbol{\theta}^{(n)})$. These are the dependent variable values for the $d \in \{1,\ldots,D\}$ regression models that we must fit. In order to generate corresponding realisations of the independent variable for the regression analysis we sample, for each value $\boldsymbol{\theta}^{(n)}$, a dataset $\mathbf{x}^{(n)}$ from the likelihood $p(\mathbf{X}|\boldsymbol{\theta} = \boldsymbol{\theta}^{(n)})$, and for each dataset calculate the summary statistic $T(\mathbf{x}^{(n)})$.

### 4.3 Choice of summary statistic $T(\mathbf{x})$

If we expect that the study data $\mathbf{x}$ to be informative for a single economic model parameter $\theta_i$, and if $\mathbf{x}$ is scalar we choose $T(\mathbf{x}) = \mathbf{x}$. If $\mathbf{x}$ is vector valued, then we choose $T(\mathbf{x})$ to be a scalar sample estimator for $\theta_i$. This leads to quite natural summary statistics. So, for example, if we wish to calculate the expected value of a two-arm, binary outcome trial to update beliefs about an odds ratio, then we our choice of $T(\mathbf{x})$ would be the sample odds ratio.

If we wish to update beliefs about $p > 1$ economic model parameters $\{\theta_1, \ldots, \theta_p\}$, then we would calculate $p$ scalar summary statistics $T(\mathbf{x}) = \{T_1(\mathbf{x}), \ldots, T_p(\mathbf{x})\}$, where each $T_i(\mathbf{x})$ is the sample estimator for $\theta_i$. For example, if we wish to calculate the expected value of a study to learn about the shape and scale parameters of a Weibull distribution, $\{\theta_1, \theta_2\}$, and $\mathbf{x}$ are censored time-to-event data, then we would choose $\{T_1(\mathbf{x}), T_2(\mathbf{x})\}$ to be the sample estimates $\hat{\boldsymbol{\theta}}_i = \{\hat{\theta}_1, \hat{\theta}_2\}$ derived from a Weibull survival model.

In the case of vector valued $T(\mathbf{x}) = \{T_1(\mathbf{x}), \ldots, T_p(\mathbf{x})\}$ we fit the multivariate model,

$$\text{NB}(d, \boldsymbol{\theta}^{(n)}) = g_1\{d, T_1(\mathbf{x}^{(n)})\} + \ldots + g_p\{d, T_p(\mathbf{x}^{(n)})\} + \varepsilon^{(n)}. \tag{22}$$

#### 4.3.1 EVSI calculation

After fitting a GAM model for each decision option $d \in \{1, \ldots, D\}$ we extract the regression model fitted values. The fitted values are estimates of $g\{d, T(\mathbf{x}^{(1)})\}, \ldots, g\{d, T(\mathbf{x}^{(N)})\}$, our target quantity. We denote the fitted values for decision option $d$ as $\hat{g}_d^{(1)}, \ldots, \hat{g}_d^{(N)}$ and the estimated EVSI is then given by

$$\widehat{\text{EVSI}}(\mathbf{X}) = \frac{1}{N}\sum_{n=1}^{N} \max_d \hat{g}_d^{(n)} - \max_d \frac{1}{N}\sum_{n=1}^{N} \hat{g}_d^{(n)}. \tag{23}$$

The sampling scheme for the GAM regression-based EVSI is given in Box 2.

**Box 2 - GAM regression-based EVSI algorithm**

*Generate a PSA sample of size $N$:*

for $k = 1, \ldots, N$ do

Sample $\boldsymbol{\theta}^{(n)}$ from the distribution of the parameters, $p(\boldsymbol{\theta})$

Evaluate the economic model to obtain (incremental) net benefits $\text{NB}(d, \boldsymbol{\theta}^{(n)})$

end for

*Given the PSA sample, simulate data samples:*

for $n = 1, \ldots, N$ do

Generate a data sample $\mathbf{x}^{(n)}$ from $p(\mathbf{X}|\boldsymbol{\theta}^{(n)})$

Calculate summary statistic $T(\mathbf{x}^{(n)})$

end for

*Fit regression models and calculate EVSI:*

Regress net benefits $\text{NB}(d, \boldsymbol{\theta}^{(n)})$ from the PSA on $T(\mathbf{x}^{(n)})$ for each $d$

Extract GAM model fitted values for each $d$

Calculate EVSI via equation (19).

# 5 Case studies

***In this section we illustrate the method. How does the regression method perform when compared with 2-level Monte Carlo?***

## 5.1 Case study 1: partial EVPI

Case study 1 is based on a hybrid decision tree / Markov model first used for illustrative purposes in Brennan et al. (2007). We very briefly describe it here. The model predicts net benefit, $\text{NB}(d, \boldsymbol{\theta})$, under two competing drug treatment options ($d = 1, 2$) and can be written in sum product form as

$$\text{NB}(1, \boldsymbol{\theta}) = \lambda \left\{ \sum_{i=1}^{20} \left( \mathbf{S}_1^{\mathrm{T}} \mathbf{M}_1^i \mathbf{U}_1 \right) + \theta_7 \theta_8 \theta_9 \right\} - (\theta_1 + \theta_2 \theta_3 \theta_4), \tag{24}$$

$$\text{NB}(2, \boldsymbol{\theta}) = \lambda \left\{ \sum_{i=1}^{20} \left( \mathbf{S}_2^{\mathrm{T}} \mathbf{M}_2^i \mathbf{U}_2 \right) + \theta_{15} \theta_{16} \theta_{17} \right\} - (\theta_{10} + \theta_{11} \theta_{12} \theta_4). \tag{25}$$

Three health states are defined, 'responding to drug', 'not-responding to drug' and 'dead'. The vectors $\mathbf{S}_1 = (\theta_5, 1-\theta_5, 0)^{\mathrm{T}}$ and $\mathbf{S}_2 = (\theta_{13}, 1-\theta_{13}, 0)^{\mathrm{T}}$ represent the proportion of the population in each health state under each decision option at time point 0. The vectors $\mathbf{U}_1 = (\theta_6, 0, 0)^{\mathrm{T}}$ and $\mathbf{U}_2 = (\theta_{14}, 0, 0)^{\mathrm{T}}$ represent the utility weights associated with each health state. The proportions in each health state are updated on a yearly cycle via the transition matrices

$$\mathbf{M}_1 = \begin{pmatrix} \theta_{18} & \theta_{19} & \theta_{20} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{M}_2 = \begin{pmatrix} \theta_{24} & \theta_{25} & \theta_{26} \\ \theta_{27} & \theta_{28} & \theta_{29} \\ 0 & 0 & 1 \end{pmatrix}. \tag{26}$$

Distributions for the transition matrix parameters are Dirichlet with $(\theta_{18-20}) \sim$ D(70,40,10); $(\theta_{21-23}) \sim$ D(10,100,20); $(\theta_{24-26}) \sim$ D(70,40,10) and $(\theta_{27-29}) \sim$ D(10,100,20). Means and standard deviations for the remaining input parameters are shown in Table 1. We expressed our uncertainty about $\theta_{2,5,7,11,13,15}$ using Beta distributions, and our uncertainty about $\theta_{3,4,9,12,17}$ using Gamma distributions.

| Parameter | Mean (sd) | |
|---|---|---|
| | $d = 1$ | $d = 2$ |
| Cost of Drug ($\theta_1, \theta_{10}$) | £1000 (£1) | £1500 (£1) |
| % Admissions ($\theta_2, \theta_{11}$) | 10% (2%) | 8% (2%) |
| Days in Hospital ($\theta_3, \theta_{12}$) | 5.20 (1.00) | 6.10 (1.00) |
| Cost per day ($\theta_4$) | £400 (£200) | £400 (£200) |
| % Responding to drug ($\theta_5, \theta_{13}$) | 70% (10%) | 80% (10%) |
| Increment in utility if respond ($\theta_6, \theta_{14}$) | 0.30 (0.10) | 0.30 (0.05) |
| % Side effects ($\theta_7, \theta_{15}$) | 25% (10%) | 20% (5%) |
| Decrement in utility if side effect ($\theta_8, \theta_{16}$) | 0.10 (0.02) | 0.10 (0.02) |
| Duration of side effect (years) ($\theta_9, \theta_{17}$) | 0.50 (0.20) | 0.50 (0.20) |

Table 1: Summary of means and standard deviations for case study model parameters

The willingness to pay for one unit of health output in QALYs is $\lambda = £10,000$/QALY. Note that some components of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{19})$ are redundant in $\text{NB}(d, \boldsymbol{\theta})$ for each $d$.

We defined three parameter sets of interest: set 1 comprising effectiveness parameters $\theta_5$ and $\theta_{13}$, representing information that could be gained from a trial; set 2 comprising effectiveness

and utility parameters $\theta_5, \theta_6, \theta_{13}$ and $\theta_{14}$, representing information that could be gained from a trial that also collected utility data; and set 3 comprising the transition matrix parameters, representing information that could be gained from the long term follow up of trial participants.

Although the case study model is computationally cheap to evaluate, we assume that we may be in a position of only being able to evaluate the model 10,000 times. We calculated partial EVPI using three approaches. Firstly, we calculated the partial EVPI for each parameter set using the two-level Monte Carlo approach with $K = 10^4$ and $J = 10^3$ and hence $10^7$ model runs. We considered this the 'gold standard'. Next, we calculated the partial EVPI values using the two-level Monte Carlo approach with three different sets of $J$ inner loop samples and $K$ outer loop samples, where $J \times K = 10,000$ model runs in total (see Table 4). Thirdly, we computed the partial EVPI values using the GAM regression method with a total of $10,000$ PSA samples. Estimates of partial EVPI using the two-level Monte Carlo method are upwardly biased for small values of $J$, due to the maximisation step. The estimates of upward bias were obtained using the method presented in Oakley et al. (2010). Standard error calculations are given in Strong et al. (2013b). For each method we report the mean CPU time taken to compute the partial EVPI for the three parameters sets of interest.

### 5.1.1 Results for the partial EVPI analysis

Table 2 shows the estimated partial EVPI values. The overall EVPI is £775. Standard errors for the 'gold standard' 2-level Monte Carlo estimates with $10^7$ model runs are small, as are the values of the upward bias. When the number of model evaluations is restricted to $10^4$ the regression methods perform considerably better than the two level Monte Carlo method, resulting in estimates that have both minimal upward bias *and* substantially greater precision. With a PSA sample size of $10^4$ the GAM takes approximately 1 second whereas, the two-level Monte Carlo method with $10^7$ model runs takes 1.8 hours.

| Sample sizes | | | Partial EVPI (SE; *upward bias*), £ | | | Mean CPU time |
|---|---|---|---|---|---|---|
| Outer loop | Inner loop | Total | Parameter set 1 $\{\theta_5, \theta_{13}\}$ | Parameter set 2 $\{\theta_5, \theta_6, \theta_{13}, \theta_{14}\}$ | Parameter set 3 $\{\theta_{18}$ to $\theta_{29}\}$ | |
| *Two-level Monte Carlo ('Gold Standard')*[a] | | | | | | |
| $10^4$ | $10^3$ | $10^7$ | 68.84 (4.47; *0.22*) | 595.14 (9.39; *0.13*) | 426.67 (6.51; *0.30*) | 1.8 hours |
| *Two-level Monte Carlo*[b] | | | | | | |
| $10^1$ | $10^3$ | $10^4$ | 5.77 (50.66; *2.01*) | 389.93 (296.6; *0.29*) | 178.93 (223.02; *0.39*) | 3.7 sec |
| $10^2$ | $10^2$ | $10^4$ | 77.07 (21.96; *19.98*) | 661.75 (94.34; *2.40*) | 362.82 (71.78; *4.41*) | 2.9 sec |
| $10^3$ | $10^1$ | $10^4$ | 228.32 (15.11; *148.24*) | 623.70 (31.43; *21.63*) | 467.61 (26.06; *42.32*) | 2.8 sec |
| *GAM regression*[b] | | | | | | |
| $10^4$ | - | $10^4$ | 62.53 (9.98; *0.47*) | 582.03 (8.23; *0.49*) | 409.80 (10.37; *1.03*) | 0.9 sec |

a. $J$ and $K$ chosen to achieve SE and bias of the same order of magnitude as the regression estimates
b. Model runs restricted to $10^4$

Table 2: Partial EVPI values and timings for case study 2

## 5.2 Case study 2 - Ades 2004 decision tree model

Our case study is based on the model that was used for illustrative purposes in Ades et al. (2004). The decision problem has two options: $d = 1$ "standard care" and $d = 2$ "new treatment", and can be represented by a simple decision tree (figure 2). There are 11 parameters in the model, which we write as the vector $\boldsymbol{\theta} = (L, Q_E, Q_{SE}, C_E, C_T, C_{SE}, P_C, P_{SE}, OR, P_T, \lambda)$. Parameter definitions and distributions are given in table 3. The output of the model is the net benefit for

each decision option in monetary units. The algebraic form of the model is given in equations (27) and (28), with some components of $\boldsymbol{\theta}$ being redundant in each net benefit function.

$$\begin{aligned} \text{NB}(1, \boldsymbol{\theta}) &= P_C \, (\lambda \, L \, (1 + Q_E)/2 - C_E) + (1 - P_C) \, \lambda \, L. \qquad (27) \\ \text{NB}(2, \boldsymbol{\theta}) &= P_{SE} \, P_T \, (\lambda \, (L \, (1 + Q_E)/2 - Q_{SE}) - (C_T + C_{SE} + C_E)) + \\ &\quad P_{SE} \, (1 - P_T) \, (\lambda \, (L - Q_{SE}) - (C_T + C_{SE})) + \\ &\quad (1 - P_{SE}) \, P_T \, (\lambda \, L \, (1 + Q_E)/2 - (C_T + C_E)) + \\ &\quad (1 - P_{SE}) \, (1 - P_T) \, (\lambda \, L - C_T). \qquad (28) \end{aligned}$$

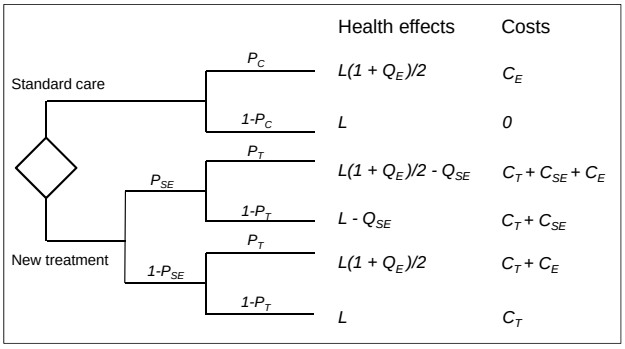


Figure 2: Decision tree model from Ades et al. (2004)

| Description | Parameter | Mean | Distribution |
|---|---|---|---|
| Mean remaining lifetime | $L$ | 30 | Constant |
| QALY after critical event, per year | $Q_E$ | 0.6405 | $\text{logit}(Q_E) \sim N(0.6, 1/6)$ |
| QALY decrement due to side effects | $Q_{SE}$ | 1 | Constant |
| Cost of critical event | $C_E$ | \$200,000 | Constant |
| Cost of treatment | $C_T$ | \$15,000 | Constant |
| Cost of treatment side effects | $C_{SE}$ | \$100,000 | Constant |
| Probability of critical event, no treatment | $P_C$ | 0.15 | Beta(15,85) |
| Probability of treatment side effects | $P_{SE}$ | 0.25 | Beta(3,9) |
| Odds ratio, $(P_T/(1-P_T))/(P_C/(1-P_C))$ | $OR$ | 0.2636 | $\log(OR) \sim N(-1.5, 1/3)$ |
| Probability of critical event on treatment | $P_T$ | 0.0440 | [derived from $OR$ and $P_C$] |
| Monetary value of 1 QALY | $\lambda$ | \$75,000 | Constant |

Table 3: Parameter distributions for case study 1 (taken from Ades et al., 2004)

For our case study we consider two of the three data collection scenarios that were presented in Ades et al. (2004), i.e. data collection to inform the quality of life after critical event ($Q_E$) and data collection to inform the treatment effect size ($OR$). For each scenario we calculated EVSI using two methods. Firstly, we used the two-level Monte Carlo scheme outlined in section 3.2, with an MCMC inner loop, and secondly we used the GAM regression method presented in section 4.2.

Because each method uses Monte Carlo to estimate the outer expectation in (10) there will be a Monte Carlo sampling error that tends to zero as the outer loop sampling size increases. For each estimated EVSI we calculated the Monte Carlo standard error using the methods presented in Strong et al. (2013a). We repeated each analysis with a range of values of $K$ to demonstrate the relationship between $K$ and the Monte Carlo standard error. We also recorded the total

CPU time required to undertake the EVSI computation in order to compare the efficiency of each method at different values of $K$.

For the two-level Monte Carlo method we chose an inner loop (MCMC) sample size of $J = 10^4$ after an initial exploration to determine an adequate sample size to achieve minimal upward bias (Oakley et al., 2010). We then chose values of $K$ equal to $10^4$, $10^5$ and $10^6$. Values of $K$ greater than this required prohibitively long run times. For the GAM based method we chose values of $K$ equal to $10^4$, $10^5$ and $10^6$.

### 5.2.1 Data collection scenario 1 - EVSI for quality of life after critical event ($Q_E$)

To reduce uncertainty about $Q_E$ we considered the value of undertaking an observational study of $n = 100$ patients who have experienced a critical event. We assume that, conditional on $Q_E$, the sample mean of the logit transform of the quality of life reported in a single data collection exercise is Normally distributed with expectation $\text{logit}(Q_E)$ and variance equal to 1/50 (see Ades et al., 2004, for details).

**Method 1 - two-level nested Monte Carlo / MCMC sampling scheme:** We implemented the general two-level Monte Carlo scheme outlined in section 3.2. In an outer loop, we drew $K$ samples from the Normal(0.6, 1/6) prior distribution for $\text{logit}(Q_E)$. For each value $\text{logit}(Q_E)^{(k)}$ we generated a sample of data $\mathbf{x}^{(k)}$ from a Normal$\{\text{logit}(Q_E)^{(k)}, 1/50\}$ distribution. Conditional on each simulated trial data value $\mathbf{x}^{(k)}$ we then ran an inner loop of size $J = 10^4$. At each run of the inner loop we sampled a vector of parameter values $\theta^{(j,k)}$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x}^{(k)})$ and evaluated the model net benefit equation $\text{NB}(d, \boldsymbol{\theta}^{(j,k)})$. Finally, we calculated EVSI via equation (11).

Because $\mathbf{x}^{(k)}$ is only informative for the parameter $Q_E$, and $Q_E$ is independent of all other model parameters, drawing from $p(\boldsymbol{\theta}|\mathbf{x}^{(k)})$ reduced in this case to drawing from the posterior distribution $p(Q_E|\mathbf{x}^{(k)})$ and the prior distributions for the remaining parameters. The posterior distribution $p(Q_E|\mathbf{x}^{(k)})$ does not have a standard form and we therefore sampled in the inner loop using MCMC implemented in OpenBUGS (Lunn et al., 2009). At each inner loop step we discarded the first 1,000 MCMC samples as a burn-in.

**Method 2 - GAM regression:** We implemented the GAM regression scheme outlined in section 4.2. Firstly, we generated a PSA sample of size $K$. We calculated the incremental net benefit for each PSA sample. Next, for each parameter vector $\boldsymbol{\theta}^{(k)}$ in the PSA sample we generated a sample of data $\mathbf{x}^{(k)}$ from a Normal$\{\text{logit}(Q_E)^{(k)}, 1/50\}$ distribution. Since $\mathbf{x}^{(k)}$ is scalar we set $T(\mathbf{x}^{(k)}) = \mathbf{x}^{(k)}$. We regressed the incremental net benefits on $T(\mathbf{x}^{(k)})$, extracted the model fitted values and estimated the EVSI via equation (19).

### 5.2.2 Data collection scenario 2 - EVSI for the treatment effect size ($OR$)

To reduce uncertainty about the treatment effect size parameter $OR$ we consider the value of undertaking a randomised controlled trial with $n = 200$ patients allocated to the new treatment, and $n = 200$ patients allocated to standard care. We assume that, conditional on $P_C$ and $P_T$, where $\text{logit}(P_T) = \text{logit}(P_C) + \log(OR)$, the number of treatment successes in the new treatment group is $\mathbf{x}_T \sim \text{Binomial}(P_T, 200)$ and in the standard care group is $\mathbf{x}_C \sim \text{Binomial}(P_C, 200)$.

**Method 1 - two-level nested Monte Carlo / MCMC sampling scheme:** We implemented the general two-level Monte Carlo scheme outlined in section 3.2. In an outer loop, we drew $K$ samples from the Beta(15, 85) prior distribution for $P_C$, and $K$ samples from the Normal(-1.5, 1/3) prior distribution for $\log(OR)$. For each $k$ we calculated $\text{logit}(P_T^{(k)}) = \text{logit}(P_C^{(k)}) + \log(OR^{(k)})$. Next, for each value $P_C^{(k)}$ we generated a sample of data $\mathbf{x}_C^{(k)}$ from a Binomial($P_C^{(k)}$, 200) distribution, and for each value $P_T^{(k)}$ a sample of data $\mathbf{x}_T^{(k)}$ from a Binomial($P_T^{(k)}$, 200) distribution. Conditional on each simulated trial data vector $\mathbf{x}^{(k)} = (\mathbf{x}_C^{(k)}, \mathbf{x}_T^{(k)})$ we then ran an inner MCMC loop of size $J = 10^4$. Because we require that $\mathbf{x}^{(k)}$ is informative for the parameter $OR$, but not for $P_C$ we sampled parameter values $P_T^{(j,k)}$ from its posterior distribution $p(P_T|\mathbf{x}^{(k)})$ and values of $P_C$ from its prior distribution. We evaluated the model at each inner loop run and calculated EVSI via equation (11). The posterior distribution $p(P_T|\mathbf{x}^{(k)})$ does not have a standard form and we therefore implemented the inner loop in OpenBUGS (Lunn et al., 2009). At each inner loop step we discarded the first 1,000 MCMC samples as a burn-in.

**Method 2 - GAM regression:** We implemented the GAM regression scheme outlined in section 4.2. Firstly, we generated a PSA sample of size $K$. We calculated the incremental net benefit for each PSA sample. Next, for each parameter vector $\boldsymbol{\theta}^{(k)}$ in the PSA sample we generated a sample of data comprising $\mathbf{x}_C^{(k)}$ from a Binomial($P_C^{(k)}$, 200) distribution, and $\mathbf{x}_T^{(k)}$ from a Binomial($P_T^{(k)}$, 200) distribution. Given data $\mathbf{x}^{(k)} = (\mathbf{x}_C^{(k)}, \mathbf{x}_T^{(k)})$ we calculated the sample log odds ratio statistic $T(\mathbf{x}^{(k)}) = \log[\{\mathbf{x}_T^{(k)}/(200 - \mathbf{x}_T^{(k)})\}/\{\mathbf{x}_C^{(k)}/(200 - \mathbf{x}_C^{(k)})\}]$. We regressed the incremental net benefits on $T(\mathbf{x}^{(k)})$, extracted the model fitted values and estimated the EVSI via equation (19).

### 5.2.3 Results for the EVSI analysis

Table 4 shows the EVSI values, standard errors and timings for the three data collection scenarios calculated by the Ades (2004) method, the two-level Monte Carlo / MCMC method, and the GAM regression method. For comparison Ades et al. (2004) reports values of \$1880 and \$3260 for scenarios 1 and 2 respectively, based on an analytic approximation to the inner loop (that is available due to the multi-linear form of this model) and an outer sample size of $10^5$.

| Sample sizes | | | EVSI (SE), \$ | | Mean CPU time |
|---|---|---|---|---|---|
| Outer ($K$) | Inner ($J$) | Total | Scenario 1 | Scenario 2 | |
| *Two-level Monte Carlo method* | | | | | |
| $10^4$ | $10^4$ | $10^8$ | 1871 (37) | 2967 (80) | 1.2 hours |
| $10^5$ | $10^4$ | $10^9$ | 1892 (12) | 3049 (26) | 12.0 hours |
| $10^6$ | $10^4$ | $10^{10}$ | 1886 (3.7) | 3031 (8.1) | 117.9 hours |
| *GAM regression method* | | | | | |
| $10^4$ | - | $10^4$ | 2047 (163) | 3117 (137) | 0.1 sec |
| $10^5$ | - | $10^5$ | 1846 (51) | 3020 (41) | 0.7 sec |
| $10^6$ | - | $10^6$ | 1861 (16) | 3035 (13) | 8 sec |

* Standard errors and timings are not reported in Ades (2004)

Table 4: Estimated EVSI values and timings for the three case study scenarios

For comparisons within each method, the Monte Carlo standard error scales in proportion to $K^{-1/2}$ as expected. Results obtained using the regression based method have slightly lower precision than those obtained by either of the other methods, but the difference is small and less than one order of magnitude. The two-level Monte Carlo method is four-five orders of magnitude slower than the the regression based method for equivalent $K$, primarily due to the requirement for $10^4$ inner loop model evaluations at each outer loop run. For both methods the computation times scale approximately in proportion to $K$.

# 6 Discussion

***What are the main findings from the case studies? What are the strengths and limitations of the regression method? What are the implications for using this method in patient-level simulation models? What are the policy implications?***

## 6.1 Main result

The regression based approach we propose requires only the single set of model evaluations that is generated in a standard probabilistic sensitivity analysis in order to calculate partial EVPI or EVSI. The regression method leads to a considerable gain in precision over the two-level Monte Carlo method with the same number of model runs, while retaining an acceptably small upward bias. The GAM method in particular is straightforward to implement in the freely available software R, thus allowing an analyst to compute VoI measures for any subset of input parameters quickly and with relative ease.

The regression method allows the complete separation of the VoI calculation step from the model evaluation step, which may be particularly useful when the model has been built using specialist software that does not allow easy implementation of the VoI step, or where those who wish to compute VoI do not 'own' (and therefore cannot directly evaluate) the model. The method has the particular advantage that it avoids the need for MCMC, a process which an analyst could find challenging without the necessary statistical training. In terms of computational speed, the regression methods are fast and we see that this will be particularly useful when the analyst is faced with a slow individual-level simulation model.

The non-parametric regression approach has some similarities to the model emulation method proposed in Oakley (2009). In one sense the GAM model can be viewed as an emulator for the inner conditional expectation in equations (7) and (11). The important difference between the two approaches is that in Oakley (2009) the net benefit function itself is emulated. Emulating the net benefit function allows for the rapid evaluation of a slow economic model, but it does not address the problem of how to sample from a difficult conditional distribution.

## 6.2 Limitations

There are some limitations of the regression approach. In general, the GAM method is straightforward to implement due to the easy availability of software (e.g. the `mgcv` package in R). However, in the case of partial EVPI estimation, if the set of parameters of interest that we wish to calculate a combined partial EVPI for is moderately large (above six or so), and if it is expected that those parameters will jointly interact (non-additively) within the economic model,

then the number of GAM model parameters that need to be estimated could exceed the number of PSA samples, causing the method to fail. In this case we would recommend using a Gaussian process regression, rather than a GAM, approach (see Strong et al., 2013b, for more details, and for accompanying R code).

In common with any method for calculating EVSI, we must define, *a priori*, a data generating process in the form of $p(\mathbf{X}|\boldsymbol{\theta})$, the density of the study data conditional on the model parameters. For studies with complex design and / or complex analysis (including studies, for example, in which observations may be censored) this requires careful thought. There is at present little research published on this topic to guide the analyst.

In order to compute EVSI, the method requires us to specify a summary statistic $T(\mathbf{x})$. If $T(\mathbf{x})$ is sufficient for $\mathbf{x}$ then the GAM method will estimate the expected value of updating the parameters conditional on all the data. If, however, $T(\mathbf{x})$ is not sufficient, then the method will instead estimate the expected value of updating the parameters conditional on $T(\mathbf{x})$. This may be undesirable, in which case we need to ensure that $T(\mathbf{x})$ is either sufficient, or as close to sufficient as possible. However, in some circumstances we may actually want to condition on $T(\mathbf{x})$ rather than all the data. If we expect that, were the proposed study actually to be done, we would only have access to summary statistics from the study (from say a report or published paper), then we should update parameters conditional on those summary statistics, not on the individual level data.

### 6.3 Using the method in patient level models

We calculate Value of Information for a patient level model via the regression method in the same manner as we would for a cohort model, i.e. by regressing the PSA sample net benefits on the parameters of interest for partial EVPI, and on the sampled data for EVSI. We briefly recap here the computation of a PSA for a patient level model. This is a two-level process whereby samples are drawn from the PSA level (i.e. population level) parameters in an outer loop, and then, conditional on these samples, individual patients are sampled in an inner loop. The purpose of sampling individual patients is to average over heterogeneity (and/or uncertainty) at the patient level for each sample of population level input parameters. 'Convergence' is achieved when the patient level sample size is large enough that, given some arbitrary sample from the PSA (population) level parameters, the estimated net benefit has sufficiently small variance. Non-convergence will introduce additional noise in the estimation of the net benefit for each sample from the PSA level parameters.

Now, recall that in our regression approach we treat *all* variability in the net benefit that is not due to the parameters of interest (in the case of EVPI) or data (in the case of EVSI) as noise. Any residual variability due to non-convergence of the patient level simulation will be treated as noise in the regression and 'averaged out'. Since the regression estimation occurs *before* the maximisation step, the residual first order variability will not cause an upward bias in the VoI estimate.

This has profound implications. O'Hagan et al. (2007) showed that if we only wish to estimate the expected net benefit for each decision option, then the optimal number of patients to run per PSA sample is 1 (assuming that patients do not interact in the model). The reason that we would usually sample more than 1 patient per PSA run is that this allows estimation of the

population expected net benefits, conditional on each PSA sample from the parameters. This then allows the CEAC and overall EVPI to be computed. However, if we adopt our regression approach, we no longer need to sample more than 1 patient at each PSA run, greatly reducing computation time. The patient-level 'first order' variability will be averaged out in the regression analysis. A regression of the net benefits on the whole set of model parameters will allow the estimation of overall EVPI, and construction of the CEAC. Regression on subsets of parameters will allow estimation of partial EVPI, and regression on simulated datasets will allow estimation of EVSI. We expect to present these important results in a full paper shortly.

### 6.4 Implications and conclusion

The key implication of our novel regression approach is that the computation of partial EVPI and EVSI have become tractable for any decision problem. With the increasing use of patient level simulation models we envisage that obtaining VoI via the traditional two-level Monte Carlo approach will be considered just too time-consuming (in fact experience suggests that the two-level Monte Carlo procedure is considered too difficult for even moderately simple cohort models). We hope that the computation of VoI now becomes standard practice, and we urge those who write guidance on good modelling practice to promote the routine reporting of VoI.

## References


Ades, A. E., Lu, G. and Claxton, K. (2004). Expected value of sample information calculations in medical decision modeling, *Medical Decision Making*, **24 (2)**: 207–227.

Brennan, A., Kharroubi, S., O'Hagan, A. and Chilcott, J. (2007). Calculating partial expected value of perfect information via Monte Carlo sampling algorithms, *Medical Decision Making*, **27 (4)**: 448–470.

Claxton, K. (1999a). Bayesian approaches to the value of information: implications for the regulation of new pharmaceuticals, *Health Economics*, **8 (3)**: 269–274.

Claxton, K. (1999b). The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies, *Journal of Health Economics*, **18 (3)**: 341–364.

Claxton, K. and Posnett, J. (1996). An economic approach to clinical trial design and research priority-setting, *Health Economics*, **5 (6)**: 513–524.

Eckermann, S. and Willan, A. R. (2007). Expected value of information and decision making in HTA, *Health Economics*, **16 (2)**: 195–209.

Felli, J. C. and Hazen, G. B. (1998). Sensitivity analysis and the expected value of perfect information, *Medical Decision Making*, **18 (1)**: 95–109.

Felli, J. C. and Hazen, G. B. (2003). Erratum: Correction: Sensitivity analysis and the expected value of perfect information, *Medical Decision Making*, **23 (1)**: 97.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models, *Statistical Science*, **1 (3)**: 297–318.

Kharroubi, S. A., Brennan, A. and Strong, M. (2011). Estimating expected value of sample information for incomplete data models using Bayesian approximation, *Medical Decision Making*, **31 (6)**: 839–852.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique, and future directions, *Statistics in Medicine*, **28 (25)**: 3049–3067.

McKenna, C. and Claxton, K. (2011). Addressing adoption and research design decisions simultaneously, *Medical Decision Making*, **31 (6)**: 853–865.

Oakley, J. E. (2009). Decision-theoretic sensitivity analysis for complex computer models, *Technometrics*, **51 (2)**: 121–129.

Oakley, J. E., Brennan, A., Tappenden, P. and Chilcott, J. (2010). Simulation sample sizes for Monte Carlo partial EVPI calculations, *Journal of Health Economics*, **29 (3)**: 468–477.

O'Hagan, A., Stevenson, M. and Madan, J. (2007). Monte carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA, *Health Economics*, **16 (10)**: 1009–1023.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Raiffa, H. (1968). *Decision Analysis. Introductory Lectures on Choices Under Uncertainty*, Reading, Massachusetts: Addison-Wesley.

Sadatsafavi, M., Bansback, N., Zafari, Z., Najafzadeh, M. and Marra, C. (2013). Need for speed: An efficient algorithm for calculation of single-parameter expected value of partial perfect information, *Value in Health*, **16 (2)**: 438–48.

Soares, M. O., Dumville, J. C., Ashby, R. L., Iglesias, C. P., Bojke, L., Adderley, U., McGinnis, E., Stubbs, N., Torgerson, D. J., Claxton, K. and Cullum, N. (2012). Methods to assess cost-effectiveness and value of further research when data are sparse: Negative-pressure wound therapy for severe pressure ulcers, *Medical Decision Making*.

Strong, M., Oakley, J. E. and Brennan, A. (2013a). An efficient method for computing the Expected Value of Sample Information. A non-parametric regression approach, *ScHARR Working Paper*, Available from `http://www.shef.ac.uk/scharr/sections/ph/staff/profiles/mark`.

Strong, M., Oakley, J. E. and Brennan, A. (2013b). Estimating multi-parameter partial Expected Value of Perfect Information from a probabilistic sensitivity analysis sample: a non-parametric regression approach, *Medical Decision Making*.

Strong, M., Oakley, J. E. and Chilcott, J. (2012). Managing structural uncertainty in health economic decision models: a discrepancy approach, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **61 (1)**: 25–45.

Welton, N. J., Madan, J. J., Caldwell, D. M., Peters, T. J. and Ades, A. E. (2013). Expected Value of Sample Information for multi-arm cluster randomized trials with binary outcomes, *Medical Decision Making*.

Wood, S. (2006). *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC.