

Efficient Computation of Partial Expected Value of Sample Information Using Bayesian Approximation

Alan Brennan*, and Samer A Kharroubi

University of Sheffield, England.

August 1, 2005

Abstract

We describe a novel process for transforming the efficiency of partial expected value of sample information (EVSI) computation in decision models. Traditional EVSI computation begins with Monte Carlo sampling to produce new simulated data-sets with a specified sample size. Each simulated new data-set is synthesised with existing prior information via Bayesian updating giving posterior probability distributions for model parameters and then a further 'inner level' Monte Carlo sampling quantifies the effect of the simulated data on the decision. This paper describes the development of a novel form of Bayesian Laplace approximation, which can be used to replace the inner Monte Carlo sampling. We illustrate the 1st and 2nd order versions of the Laplace approximation formula, comparing EVSI estimates in two case study cost-effectiveness models. Computation time reductions are up to 150 times shorter in our case studies and of considerably higher order for more complex models. Computational efficiency gains depend on the complexity of the net benefit functions, the number of inner level Monte-Carlo samples used, and the requirement or otherwise for Markov Chain Monte Carlo (MCMC) methods to produce the Bayesian updated posterior distributions. This methodology provides a new and valuable approach for EVSI computation in health economic decision models and potential wider benefits in many fields requiring Bayesian approximation.

*Correspondence to: Alan Brennan, Director of Health Economics and Decision Science, School of Health and Related Research (SchARR), University of Sheffield, Sheffield, South Yorkshire, S1 4DA, UK. Email: a.brennan@sheffield.ac.uk Phone: +44 (0) 114 222 0684. Fax : +44 (0) 114 222 0795.

Keywords: Bayesian decision theory; Value of information; Uncertainty; Sensitivity analysis; Sample size; Clinical trial design; Stochastic CEA; approximate Bayesian inference

JEL classification: I18

Acknowledgements: The authors would particularly like to thank Professor Tony O'Hagan for comments on this work, and Professor Trevor Sweeting for his support.

1 Introduction

1.1 Expected Value of Sample Information

Global investment in biomedical research is estimated at around \$ 80 billion at 2001/2 prices (Lewison et al, 2004). Investment by governmental bodies, pharmaceutical companies and charities aims to develop and test health technologies and to have an effect on healthcare investment decisions. The efficiency of this research investment is therefore an important issue.

Expected Value of Sample Information (EVSI) was developed within decision theory (Raiffa *et al*, 1967). The central concept of EVSI is to quantify the expected value to the decision maker which might be gained by obtaining sample information before making a large decision (e.g. test drilling for oil before setting up a major oil platform). In health economics, the use of Value of Information methods in general is an active area of methodological development, with authors investigating and promoting their use in sensitivity analysis and to quantify the value of research (Claxton *et al*, 1996 and 1999, Felli and Hazen 1998, Meltzer 2001, Brennan *et al* 2002a, Coyle *et al* 2003, Tappenden *et al* 2004). EVSI in particular is under discussion as a tool for quantifying the societal value of expensive medical research projects and for determining optimum sample sizes and allocation rates in randomized clinical trials and other research studies (Claxton *et al* 2001, Chilcott *et al* 2003, Brennan *et al* 2002a,b, and Ades *et al*, 2004).

We begin with a mathematical description of EVSI using the context of health economics (Brennan *et al* 2002b, Ades *et al*, 2004). We assume a decision model with uncertain parameters θ , for which we have a multi-variate probability distribution based on current evidence $p(\theta)$. There is a decision to be made between a fixed number of treatment strategies $t = 1, 2, \dots, T$. Each treatment delivers expected utility gains measured in quality adjusted life years (QALYs), with an associated expected cost. We adopt a net benefit approach to the cost-effectiveness analysis (Stinnett *et al*, 1998), whereby health gains are monetarised by multiplying by the decision maker's willingness to pay per additional QALY (λ). $NB(t, \theta)$ is the net benefit of treatment t if the parameters take the value θ . The net benefit for treatment t is thus $NB(t, \theta) = \lambda U(t, \theta) - C(t, \theta)$ where the functions C and U give the costs and the QALYs under treatment t .

The optimal decision given the current information is the decision that yields the highest expected net benefit. Using expectation notation (where $E_z[f(z)] = \int f(z)p(z)dz$),

we write this highest expected net benefit as,

$$\max_t E_\theta NB(t, \theta)$$

Assume that a particular research study is being considered. The new study will provide new data relating to a subset of parameters of interest θ_I . The data obtained from the proposed study X_{θ_I} will update our knowledge concerning the parameters of interest θ_I , and if parameters are correlated, may additionally tell us something the complementary set of parameters θ_I^c i.e. $\theta = (\theta_I, \theta_I^c)$. It is this process of updating the probability distributions $p(\theta)$ given new data X_{θ_I} that makes EVSI inherently Bayesian.

If we imagine we have obtained data X_{θ_I} , then we will be able to make a revised decision based on the new information. This will choose the treatment with the highest expected net benefit now given the data. This expectation is taken over the joint posterior density of $\theta|X_{\theta_I}$ and can be written as, $\max_t E_{(\theta_I^c, \theta_I|X_{\theta_I})} NB(t, \theta_I, \theta_I^c)$. If θ_I and θ_I^c are independent, then this expectation is taken over the prior density of θ_I^c , and the posterior density of θ_I given X_{θ_I} .

As yet we do not know what the result of the proposed collection of data X_{θ_I} will be. Thus, to calculate the expected value of a decision made after data have been collected, we must average over the distribution of X_{θ_I} , giving $E_{X_{\theta_I}} [\max_t E_{(\theta_I^c, \theta_I|X_{\theta_I})} NB(t, \theta_I, \theta_I^c)]$. Finally, the expected value of sample information is the difference between the expected value of a decision made after data X_{θ_I} have been collected and expected value of a decision made now, with only current information, i.e.

$$EVSI = E_{X_{\theta_I}} [\max_t E_{(\theta_I^c, \theta_I|X_{\theta_I})} NB(t, \theta_I, \theta_I^c)] - \max_t E_\theta NB(t, \theta) \quad (1)$$

1.2 Computation of EVSI via 2 level Monte Carlo Sampling

To compute EVSI, we evaluate the first term of (1), which contains an inner expectation of net benefit over $\theta_I^c, \theta_I|X_{\theta_I}$ and an outer expectation over X_{θ_I} . Current computation methods involve a two level Monte-Carlo sampling algorithm.

Box 1: 2 Level Monte Carlo Sampling Algorithm to Compute EVSI

0. Define a proposed data collection e.g. proposed sample size for data the parameters of interest

1. Begin an outer loop to simulate the variety of possible results from the proposed data collection i.e. the outer expectation.

There are two sources of variability in the possible results. The first is the uncertainty about the true underlying value of the parameters of interest. The second is that, even given the true underlying value of the parameter of interest, there is random chance associated with data collection of a specific finite sample size. Both need to be accounted for in the simulation. Sample the data collection as follows: (a) sample the true underlying values for parameters of interest θ_I from their prior distribution (e.g. sample the true value for % response rate to a drug, say $\theta_{sample1} = 60\%$) (b) sample simulated data X_{θ_I} given the sampled true underlying values of parameters of interest (e.g. sample the responses in a trial of $n=100$, from a Binomial(100,60%) given $\theta_{sample1} = 60\%$).

2. Synthesise existing evidence with simulated data.

Combine the prior knowledge with the new simulated data using Bayesian updating techniques. If the likelihood for the proposed data X_{θ_I} is conjugate with the prior distribution, then analytic formulae for the updated means or other parameters for posterior distributions are available. Examples include the normal, Beta and Gamma distributions. In the absence of conjugate distributions, Bayesian updating would generally be undertaken for using Markov Chain Monte Carlo methods (e.g. an application in WinBUGS - <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>) to allow sampling from the posterior distribution. Typically, the posterior has a revised mean and a smaller variance than the prior probability distribution. Note that Ades et al. (2004) give algorithms for the case when the parameters of interest θ_I are independent from their complement θ_I^c . In general this may not be the case and conditional distributions will need to be used to sample the θ_I^c . The process involves sampling from the marginal distribution $p(\theta_I|X_{\theta_I})$ and then sampling from the conditional distribution $p(\theta_I^c|\theta_I)$.

3. Examine impact of simulated data on the revised decision

Evaluate the inner expectation conditional on X_{θ_I} via Monte Carlo sampling i.e. re-run the probabilistic sensitivity analysis on the decision model, with updated parameters now sampled from their posterior probability distributions. Identify the 'revised decision' and quantify the net benefit obtained.

4. Loop back to repeat steps 1 to 3 (say 10,000 times) in order to simulate across the

variety of results from the proposed data collection. The EVSI for the proposed data collection is the average of the net benefits of revised decisions found in (3) minus the average net benefit provided by the 'baseline decision' without additional data collection.

This nested evaluation of Monte-Carlo estimates produces significant computation time. Previous authors restrict their work to conjugate distributions, thus simplifying computation. Nevertheless we still require large numbers of decision model runs e.g. 10,000 outer sampled data-sets, Bayesian updates for each one, and then 10,000 inner runs of the decision model, producing in total 100 million model runs . If Bayesian updating is done with MCMC in WinBUGS, this creates a level of computation, which might be impractical in all but the simplest of decision models.

In this paper, we develop a novel refinement of previously published Bayesian approximation methods, which it transpires, can apply successfully in EVSI computation transforming the efficiency of the process (section 2). We apply the new methodology to 2 case studies comparing the results with the traditional 2 level approach, and examine the computation time required to achieve comparable results (section 3). Finally, we discuss the implications and potential further use of the approach (section 4). Appendix 1 sets out the derivation of the key Bayesian approximation formula.

2 Bayesian Approximation Methodology

2.1 Existing Approximation Formulae Using the Signed Root Log-likelihood Ratio (SRLLR)

We begin by reviewing the key elements of recent work on Bayesian approximation by Sweeting and Kharroubi (2003). This work developed new accurate approximations for a number of Bayesian quantities. From the perspective of EVSI, by far the most important of these is an approximation formula for the posterior expectation of a real valued function of several parameters given particular sample data. We present the necessary notation and give the result; the reader is referred to Sweeting and Kharroubi (2003) for the detailed derivation.

Suppose that the observed data X consist of n independent and identically distributed observations. Let $p(X|\theta)$ be the probability density of X given a vector of d parameters $\theta = (\theta^1, \theta^2, \dots, \theta^d)$ i.e. the likelihood $L(\theta)$ of the observed data. Suppose that a prior

probability density $\lambda(\theta)$ for θ is available and denote by $\pi(\theta|X)$ the associated posterior density of θ given X .

Write $\theta = (\theta^1, \dots, \theta^d)$ and let $l(\theta) = \log L(\theta)$, $l_i(\theta) = \partial l(\theta)/\partial \theta^i$ and $j(\theta) = -d^2 l/d\theta^2$, the matrix of second-order partial derivatives of $-l(\theta)$. Assume that the maximum likelihood (ML) estimate $\hat{\theta}$ exists and write $J = j(\hat{\theta})$. Let $\theta_i = (\theta^1, \dots, \theta^i)$ and $\theta^{(i)} = (\theta^i, \dots, \theta^d)$, the vectors of the first i and last $(d - i + 1)$ components of θ . Define $\hat{\theta}^{(i)}(\theta_{i-1})$ to be the ML estimate of $\theta^{(i)}$ conditional on θ_{i-1} and, for $j \geq i$, let $\hat{\theta}^j(\theta_{i-1})$ be the j th component of $(\theta_{i-1}, \hat{\theta}^{(i)}(\theta_{i-1}))$. For any function $g(\theta)$, when $i < d$ we use the short-hand $g(\theta_i)$ to denote $g(\theta_i, \hat{\theta}^{(i+1)}(\theta_i))$. We define $j^{(i)}(\theta)$ to be the submatrix of $j(\theta)$ corresponding to $\theta^{(i)}$, setting $|j^{(d+1)}(\theta)| = 1$, and then define $\nu_i(\theta) = \lambda(\theta)|j^{(i+1)}(\theta)|^{-1/2}$. Finally, define the *signed-root loglikelihood ratio* transformation

$$r^i(\theta_i) = \text{sign}\{\theta^i - \hat{\theta}^i(\theta_{i-1})\} [2\{l(\theta_{i-1}) - l(\theta_i)\}]^{1/2}, \quad i = 1, \dots, d. \quad (2)$$

Now, for $i = 1, \dots, d$, define the scalars θ^{i-} and θ^{i+} as

$$\theta^{i-} = \hat{\theta}^i - (k^i)^{-1/2}, \quad \theta^{i+} = \hat{\theta}^i + (k^i)^{-1/2}, \quad (3)$$

where k^i is the reciprocal of the first entry in $(J^{(i)})^{-1}$ and write

$$\theta_i^- = (\hat{\theta}_{i-1}, \theta^{i-}, \hat{\theta}^{(i+1)}(\hat{\theta}_{i-1}, \theta^{i-}))$$

$$\theta_i^+ = (\hat{\theta}_{i-1}, \theta^{i+}, \hat{\theta}^{(i+1)}(\hat{\theta}_{i-1}, \theta^{i+})).$$

Sweeting and Kharroubi (2003) show that we can approximate the posterior expectation with the formula

$$E\{v(\theta)|X\} \cong v(\hat{\theta}) + \sum_{i=1}^d \left(\alpha_i^- v(\theta_i^-) + \alpha_i^+ v(\theta_i^+) - v(\hat{\theta}) \right). \quad (4)$$

where

$$\alpha_i^- = (\tau^i)^{-1} \{ \nu_i(\theta_i^-) / l_i(\theta_i^-) \},$$

$$\alpha_i^+ = (\tau^i)^{-1} \{ -\nu_i(\theta_i^+) / l_i(\theta_i^+) \}$$

and $\tau^i = \{ \nu_i(\theta_i^-) / l_i(\theta_i^-) \} + \{ -\nu_i(\theta_i^+) / l_i(\theta_i^+) \}$. The formulation (4) has a first term that evaluates the function $v(\theta)$ at the point where θ equals $\hat{\theta}$, the maximum likelihood estimator. From the perspective of EVSI, this means that the first term in the approximation ignores the prior probability distribution entirely, meaning that the 2nd correction term, which requires some considerable computation, will always be necessary. Within the summation, the second term evaluates the function $v(\theta)$ at 3 points θ_i^+ , θ_i^- and $\hat{\theta}$, with weights α_i^+ and α_i^- applied to θ_i^+ and θ_i^- respectively.

2.2 A Novel Refinement based on Signed Root Log-Density Ratios (SRLDR)

In this section we follow on from Sweeting and Kharroubi (2003) to develop a novel formulation for the posterior expectation approximation. Our new formulation changes the earlier approach in one fundamental way. It is designed to absorb the prior into the likelihood function, constructing formulae that focus on the posterior, which turns out to be an attractive property for various practical implications of interest, including the computation of EVSI. We obtain a variant of formula (4) for posterior expectations that is based on the signed root log-density ratio as opposed to the SRLLR.

To begin with, we regard $L(\theta)$ as a density with respect to $\lambda(\theta)d\theta$; that is, we redefine $L(\theta)$ now as $L(\theta) = \lambda(\theta)L(\theta)$. The result of this is that $L(\theta)$ can now be considered the posterior probability density and $\hat{\theta}$ can be considered the maximising θ for that posterior density i.e. the posterior mode.

From the notation of Section 2.1, define the *signed-root logdensity ratio* transformation

$$r^i(\theta_i) = \text{sign}\{\theta^i - \hat{\theta}^i(\theta_{i-1})\}[2\{l(\theta_{i-1}) - l(\theta_i)\}]^{1/2}, \quad i = 1, \dots, d. \quad (5)$$

The detailed derivation, given in *Appendix 1* produces the following alternative posterior expectation to (4):

$$E\{v(\theta)|X\} \cong \hat{v}(\hat{\theta}) + \sum_{i=1}^d \left(\alpha_i^- v(\theta_i^-) + \alpha_i^+ v(\theta_i^+) - v(\hat{\theta}) \right), \quad (6)$$

where

$$\begin{aligned} \theta_i^- &= (\hat{\theta}_{i-1}, \theta^{i-}, \hat{\theta}^{(i+1)}(\hat{\theta}_{i-1}, \theta^{i-})), \\ \theta_i^+ &= (\hat{\theta}_{i-1}, \theta^{i+}, \hat{\theta}^{(i+1)}(\hat{\theta}_{i-1}, \theta^{i+})), \\ \nu_i(\theta) &= |j^{(i+1)}(\theta)|^{-1/2}, \\ \alpha_i^- &= (\tau^i)^{-1} \{\nu_i(\theta_i^-)/l_i(\theta_i^-)\}, \\ \alpha_i^+ &= (\tau^i)^{-1} \{-\nu_i(\theta_i^+)/l_i(\theta_i^+)\}, \text{ and} \\ \tau^i &= \{\nu_i(\theta_i^-)/l_i(\theta_i^-)\} + \{-\nu_i(\theta_i^+)/l_i(\theta_i^+)\}. \end{aligned}$$

As in Section 2.1, formula (6) exhibits the posterior expectation of $v(\theta)$ as a first term $v(\hat{\theta})$, plus a correction term which evaluates the function $v(\theta)$ at three points θ_i^+ , θ_i^- and $\hat{\theta}$, with weights α_i^+ and α_i^- which do not depend on the function v at all. In

contrast to formula (4), the first term of (6) takes account of both the prior probability distribution and the collected sample data since it is expressed in terms of the posterior mode. Conceptually, this raises the possibility that the 1st term alone might provide an adequate estimation of the expectation $E\{v(\theta)|X\}$.

2.3 Conceptualising the θ^+ , θ^- , α^+ and α^-

A brief explanation of the meaning of each element of equation (6) is worthwhile. The elements are easier to begin to conceptualise in the univariate case i.e. where there is only one uncertain parameter, giving the formula,

$$E\{v(\theta)|X\} \doteq \hat{v}(\hat{\theta}) + \{\alpha^- v(\theta^-) + \alpha^+ v(\theta^+) - v(\hat{\theta})\}, \quad (7)$$

In the univariate case, θ^+ and θ^- are each simply one standard deviation away from the mode of the posterior probability density $\theta \pm \sigma$, where $\sigma = j^{-1/2}$ quantifies the standard deviation.

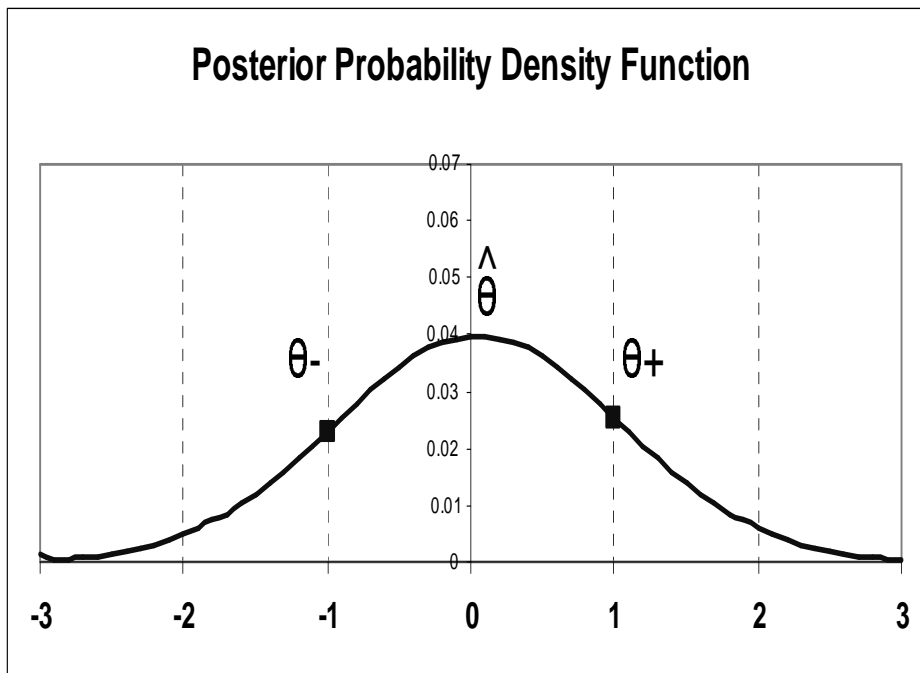


Figure 1. Illustration of $\hat{\theta}$, θ^+ and θ^- when the posterior probability density function is a normal distribution.

The approximation requires us to evaluate the function $v(\theta)$ at just these three points $\hat{\theta}$, θ^+ and θ^- . This contrasts with the evaluation of an expectation using Monte Carlo

sampling from many random points across the posterior density of θ . A weighted average of the two evaluations at θ^+ and θ^- is taken. α^+ and α^- are the weights given. In the univariate case, each is a function of the first derivative of the log of the posterior density function evaluated at both points θ^+ and θ^- . That is,

$$\alpha^+ = \frac{1}{1 - \frac{l'(\theta^+)}{l'(\theta^-)}} \quad \text{and} \quad \alpha^- = \frac{1}{1 - \frac{l'(\theta^-)}{l'(\theta^+)}} = 1 - \alpha^+$$

Note that α^\pm are approximately $1/2$. In the special case when the posterior density is symmetric, then θ^+ and θ^- will be equidistant from $\hat{\theta}$ and the first derivative (slope) of the log of the posterior density function at these 2 points will be equal and opposite i.e. $\frac{l'(\theta^-)}{l'(\theta^+)} = -1$. This results in $\alpha^+ = \alpha^- = 1/2$.

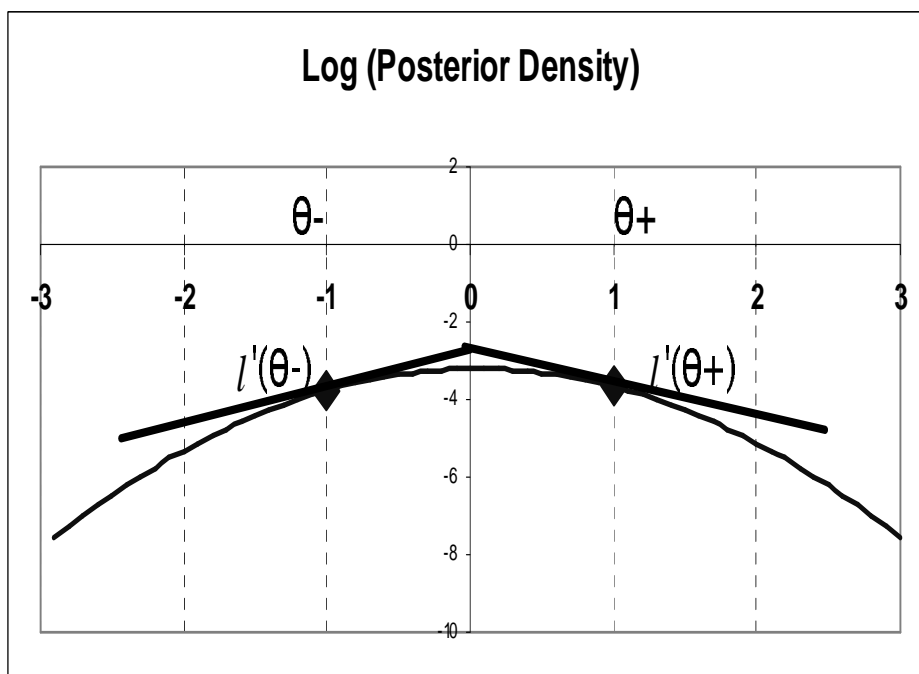


Figure 2. Illustration of α^+ and α^- when the posterior probability density function is a normal distribution.

More generally, if the posterior probability density is, say positively skewed, then the slope of the log posterior density at θ^- will be greater than at θ^+ , resulting in $\frac{l'(\theta^-)}{l'(\theta^+)}$ being greater than 1, and hence $\alpha^- < 1/2 < \alpha^+$. That is, with a positively skewed posterior distribution, more weight will be given to the evaluation of the function at $v(\theta^+)$ than to the evaluation of $v(\theta^-)$.

In evaluating the inner expectation we are effectively integrating $v(\theta)p(\theta)$. Using the 1st term of the approximation we are estimating the area under this curve by quantifying the function at posterior mode $\hat{\theta}$. This will only be accurate if the function is linear and the posterior mode is also the mean. We add the 2nd correction term which evaluates the function at the 3 points $\hat{\theta}$, θ^+ and θ^- . It is this 2nd term that accounts for any non-linearity in the function $v(\theta)$. In the example where the posterior density is normal and symmetric, such that the weights $\alpha^\pm = \frac{1}{2}$, the non linearity in the function is accounted for because we are estimating the area under the curve $v(\theta)p(\theta)$ by $\frac{1}{2}v(\theta^+) + \frac{1}{2}v(\theta^-)$.

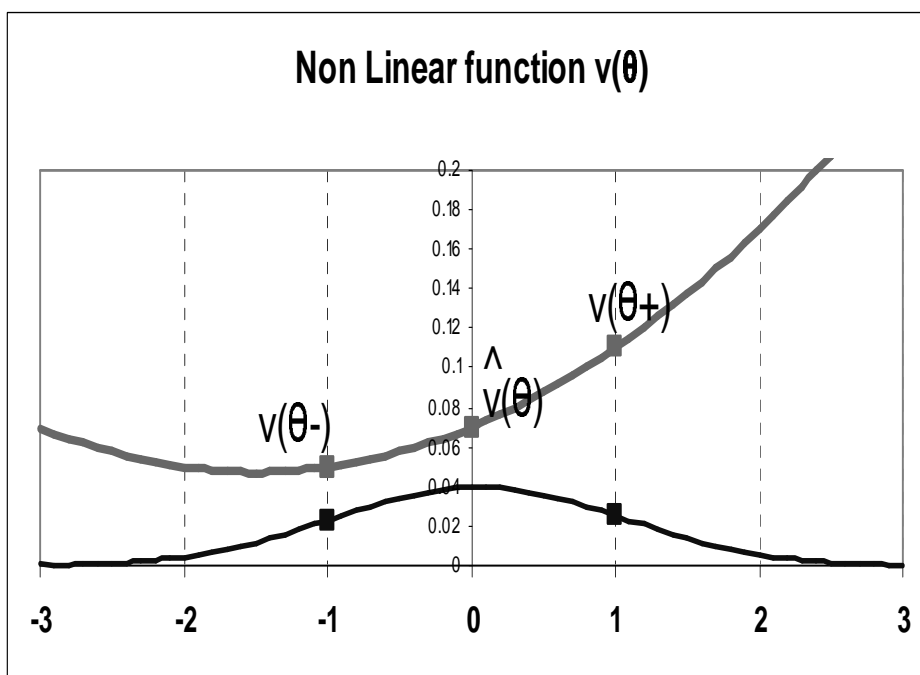


Figure 2. Illustration of the Evaluation of a Non linear function at $\hat{\theta}$, θ^+ and θ^- when the posterior probability density function is a normal distribution.

The conceptualisation of θ_i^\pm and α_i^\pm in the multi-parameter context is a complex generalisation of the univariate case. It is easier to understand their effects through the steps required in their computation.

2.4 The Steps for Computing $\hat{\theta}$, θ_i^+ , θ_i^- , α_i^+ and α_i^-

The computation of the Bayesian approximation itself contains a number of steps, which can incur reasonable time. The choice of using the approximation method versus the

traditional Monte- Carlo approach to evaluating the conditional expectation will partly depend upon this trade-off in computing time. Here, we explain how to compute each of the components of formula (6) in turn, describing a generalised process together with some short-cuts which may be available if the distributions are of a certain form.

$\hat{\theta}$ is the posterior mode i.e. it is the values $(\hat{\theta}^1, \dots, \hat{\theta}^d)$ that maximise the posterior density function given the data X . If the distributions are conjugate, then we can often use conjugate formulae to compute the posterior mode analytically. For example, if θ and X are both multi-variate normal such that $\theta \sim N(m, V)$ and $X \sim (m', V')$, then the posterior mode, which in this case is equivalent to the posterior mean, is given analytically by the formula, $\hat{\theta} = (V'^{-1} + V^{-1})^{-1}(V'^{-1}m' + V^{-1}m)$. In the general case we will need to use an iterative numerical optimisation process such as the Newton-Raphson technique to estimate $\hat{\theta}$. (In our case study applications we used [R] software with the `optim` or `nlim` functions, writing the posterior density function mathematically and then computing the $\hat{\theta}$ which minimises the negative of the posterior density function, or the negative log posterior which is often easier mathematically).

θ_i^+ and θ_i^- themselves are vectors. Each is the i th row of the matrix θ^+ and θ^- respectively. The matrix θ^+ has the following structure:

Matrix Diagram 1: θ^+

$$\theta^+ = \begin{pmatrix} \hat{\theta}^1 + (k^1)^{-\frac{1}{2}} & \leftarrow\leftarrow\leftarrow & & \hat{\theta}^{(2)}(\theta_1) & & \rightarrow\rightarrow\rightarrow & & \\ \hat{\theta}^1 & & \ddots & & & & & \\ \hat{\theta}^1 & \dots & \ddots & & & & & \\ \hat{\theta}^1 & \dots & \hat{\theta}^{i-1}, & \hat{\theta}^i + (k^i)^{-\frac{1}{2}}, & \leftarrow\leftarrow\leftarrow & \hat{\theta}^{(i+1)}(\theta_i) & \rightarrow\rightarrow\rightarrow & \\ \hat{\theta}^1 & \dots & & & \ddots & & & \\ \hat{\theta}^1 & \dots & & & & & & \\ \hat{\theta}^1 & \dots & & & & & & \hat{\theta}^d + (k^d)^{-\frac{1}{2}} \end{pmatrix}$$

The off diagonal elements, to the left of and below the diagonal, are simply the posterior modes for the first $i-1$ components of θ i.e. $\hat{\theta}^1, \hat{\theta}^2, \dots, \hat{\theta}^{i-1}$.

The diagonal elements are $\theta^{1+}, \theta^{2+}, \dots, \theta^{d+}$, where $\theta^{i+} = \hat{\theta}^i + (k^i)^{-\frac{1}{2}}$. k^i , which is a constant, comes from the information matrix J , and is the reciprocal of the 1st element of the matrix $[J^{(i)}]^{-1}$. To get k^i we need to undertake the following steps.

do need it for the first $d-1$ rows, and it needs to be done separately for θ^+ and θ^- , thus requiring $2(d-1)$ numerical optimisations.

To compute α_i^+ and α_i^- we need to compute $l_i(\theta_i^\pm)$, and $\nu_i(\theta_i^\pm) = |j^{(i+1)}(\theta_i^\pm)|^{-1/2}$. $l_i(\theta_i^+)$ is the partial derivative of the log posterior density function with respect to θ^i i.e. $\partial l(\theta)/\partial \theta^i$ evaluated at the point θ_i^+ . This is usually obtained analytically by undertaking the partial differentiation. $\nu_i(\theta_i^+)$ is obtained using the $j(\theta)$ matrix illustrated in Matrix diagram 2, and obtaining the determinant of the submatrix $j^{(i+1)}(\theta)$, when $\theta = \theta_i^+$.

Note that, in this multi-parameter situation, if the function $v(\theta)$ is linear and the parameters θ are independent and the posterior probability distribution is symmetric with the mode equal to the mean $\hat{\theta} = \theta_{mean}$, then the 1st term approximation is accurate because the matrices θ^+ and θ^- are diagonal, $\alpha^+ = \alpha^- = \frac{1}{2}$ and hence the 2nd term in the approximation is zero.

2.5 The Approximation Formula for EVSI

The inner expectation of the first term of the equation for EVSI (formula 1) can now be approximated using formula (6). We simply let $v(\theta) = NB(t, \theta)$, then the inner expectation of the first term of (1) is,

$$NB(t, \hat{\theta}) + \sum_{i=1}^d \left(\alpha_i^- NB(t, \theta_i^-) + \alpha_i^+ NB(t, \theta_i^+) - NB(t, \hat{\theta}) \right),$$

and so the Bayesian approximation formula for EVSI is, *EVSI*

$$\cong E_{X_{\theta_I}} \max_t \left\{ NB(t, \hat{\theta}) + \sum_{i=1}^d \left(\alpha_i^- NB(t, \theta_i^-) + \alpha_i^+ NB(t, \theta_i^+) - NB(t, \hat{\theta}) \right) \right\} - \max_t E_{\theta} NB(t, \theta). \quad (8)$$

Note, that the number of evaluations of the net benefit function required to estimate EVSI using the approximation formula is $3d+1$ times the number of treatment strategies under consideration.

3 Case Studies

3.1 Hypothetical Cost-Effectiveness models

To explore the feasibility of using formula (8) for the EVSI approximation, we utilised a simple hypothetical cost-effectiveness model. The cost-effectiveness model compares two strategies: treatment with drug T0 versus treatment with drug T1. Table 1 shows the nineteen uncertain model parameters, with prior mean values shown for T0 (column

a), T1 (column b) and hence the incremental analysis (column c). Costs include cost of drug and cost of hospitalisations i.e the product of the percentage of patients admitted to hospital, days in hospital, and cost per day. Thus, mean cost of strategy T0 = \$1000 + 10% x 5.20 x \$400 = \$1208. Health benefits are measured as QALYs gained and come from two sources: responders receive a utility improvement for a specified duration, and some patients have side effects with a utility decrement for a specified duration i.e. QALY for strategy T0 = 70% responders x 0.3 x 3 years + 25% side effects x -0.1 x 0.5 years = 0.6175. The willingness to pay i.e. threshold cost per QALY is set at \$10000. Thus, the net benefit of T0 is = \$10000 x 0.6175 - \$1208 = \$4967. Effectively this is a simple decision tree model with a net benefit function of sum-product form i.e.

$$NB(T0) = \lambda(\theta_5\theta_6\theta_7 + \theta_8\theta_9\theta_{10}) - (\theta_1 + \theta_2\theta_3\theta_4)$$

$$NB(T1) = \lambda(\theta_{14}\theta_{15}\theta_{16} + \theta_{17}\theta_{18}\theta_{19}) - (\theta_{11} + \theta_{12}\theta_{13}\theta_4)$$

In case study 1 the uncertain model parameters are characterised with independent normal distributions. Standard deviations for the model parameters are shown in columns (d) and (e). Each parameter can be informed by collection of further data on individual patients. It is assumed in this simple model that the patient level variance is known for each parameter and this is given in columns (f) and (g). Note, by assuming independent normal distributions with known patient level variance, we are able to use the conjugate assumption to calculate posterior densities analytically. This speeds up the process of Bayesian updating in the traditional EVSI computation algorithm because we do not need MCMC. Given current knowledge, the basic model results show \$5405 expected net benefit for T1 compared with \$4967 for T0 (difference = \$437.80), which means that our baseline decision given current information should be to adopt strategy T1. Probabilistic sensitivity analysis (Briggs *et al*, 1999) shows that T1 provides greater net benefits than T0 on only 54.5% of 1000 Monte Carlo samples. This suggests that obtaining more data on some of the uncertain parameters might help us with our decision.

In case study 2, exactly the same model is used but with correlations between several of the model parameters. The uncertainty is characterised with multi-variate normal distributions. Parameters 5,7,14 and 16 are each correlated (correlation coefficient = 0.6), and parameters 6 and 15 are independent of these but correlated with each other (again 0.6 correlation). Again because the prior distribution for the parameters and the sampling distribution for the data are assumed normal we can use analytic formulae for the Bayesian updating.

To compare EVSI methods, we investigated a series of proposed data collection exercises based on subgroups of parameters.

- a) a proposed randomized controlled clinical trial measuring only response rate parameters (parameters 5, 14)
- b) an observational study on utility only (parameters 6, 15),
- c) a trial combined with utility data collection (parameters 5, 6, 14, 15),
- d) an observational study of the duration of response to therapy (parameters 7, 16)
- e) a trial combined with utility study alongside an observational study on duration of response (parameters 5, 6, 7, 14, 15, 16).

In each proposed data collection exercise, five different sample sizes ($n= 10, 25, 50, 100$ and 200) were investigated.

The models and the programs to undertake both the traditional 2 level and new approximation EVSI computations were written in [R] and are available on the CHEBS website (<http://www.shef.ac.uk/cheps/>).

3.2 Case Study Results - Accuracy of Bayesian Approximation

For case study 1, we analysed the accuracy of the 1st order Laplace approximation i.e. using the first term in formula (8) only. Figure 4 shows the EVSI results using the 2 level Monte Carlo sampling algorithm (1,000 outer times 1,000 inner Monte Carlo samples) compared to the Bayesian approximation (150,000 simulated data-sets). Table 2 shows these EVSI results on a 100 point indexed scale relative to the overall expected value of perfect information (\$1319 per patient in our model). The differences between the two methods on this indexed scale are within 5 percentage points for all but one evaluation, which had a difference of 7 points. For practical purposes, this level of alignment in estimates means that the approximation based on only the first order term in Formula (8) and using just 1,000 simulated data-sets provides an adequate estimate of EVSI for this decision model. Note that this comparison over-estimates error differences between the approaches. This is because the results shown were independently sampled for the two methods i.e. 1,000 simulated data-sets in the 2 level traditional method , and a different set of simulated data-sets for the Laplace method. This introduces more noise than if the same sample data-sets were used in both methods and only the method for evaluating the inner integral were different.

As one would usually expect, EVSI is greater for higher sample sizes but with diminishing returns as n increases. The lower bound of EVSI for a sample size of zero is clearly zero (there is no value to be obtained from no data). The upper bound is the partial expected value of perfect information, known as partial EVPI (in a sense the value of a study with infinite sample size for a particular subgroup of parameters).

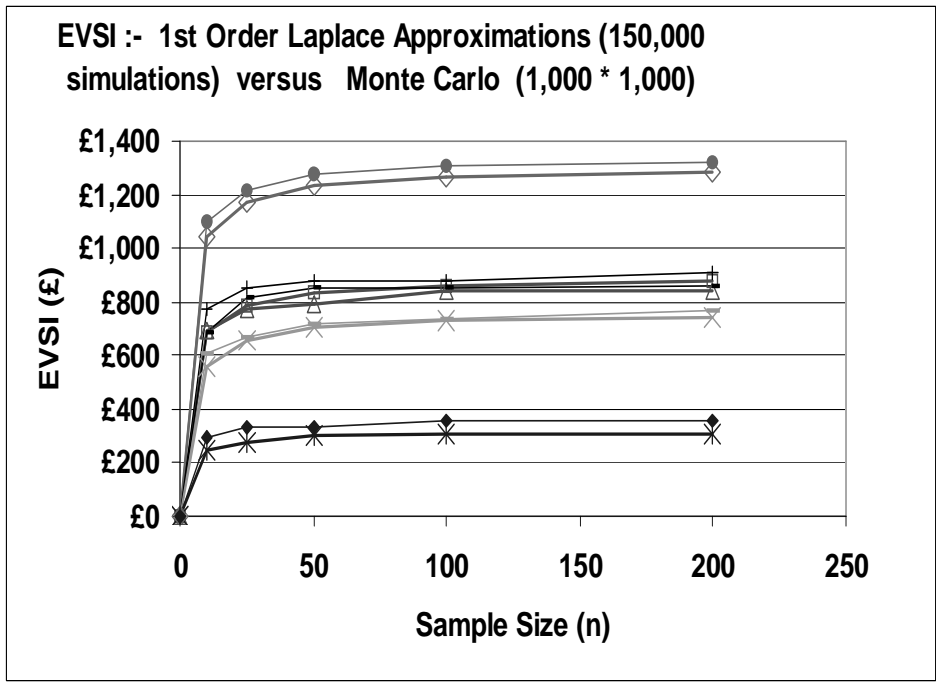


Figure 4. EVSI results for 5 proposed data collection exercises using 1st order Bayesian Approximation compared with traditional 2 level Monte Carlo simulations

Indexed EVSI Results						
	Study	Sample Size, n=				
		10	25	50	100	200
2 level Monte Carlo (1,000 * 1,000)	a	23	25	25	27	27
	b	46	51	54	56	58
	c	52	62	65	65	65
	d	59	65	67	66	69
	e	83	92	97	99	100
Bayesian Approx'n (Laplace)	a	18	21	23	23	23
	b	42	50	54	55	56
	c	53	59	60	64	64
	d	52	60	63	65	66
	e	79	89	93	96	97
Difference	a	4	4	2	4	4
	b	4	1	1	0	2
	c	-0	3	5	1	1
	d	7	5	3	1	2
	e	4	3	3	3	3

Table 2. EVSI results for 5 proposed data collection exercises in Case Study 1 using 1st order Bayesian Approximation compared with traditional 2 level Monte Carlo simulations (indexed to overall EVPI=100)

In case study 2 however, the results of the 1st order approximation were not an accurate approximation. Figure 5 shows that for a proposed study on parameters θ_6 and θ_{15} , the 1st order Laplace EVSI approximations using 10,000 simulated data-sets substantially underestimate the values obtained via a 2 level Monte Carlo run with 10,000 data-sets and 1,000 inner level simulations. When the 2nd order approximation term is used however, the accuracy more than adequate from a decision maker's perspective.

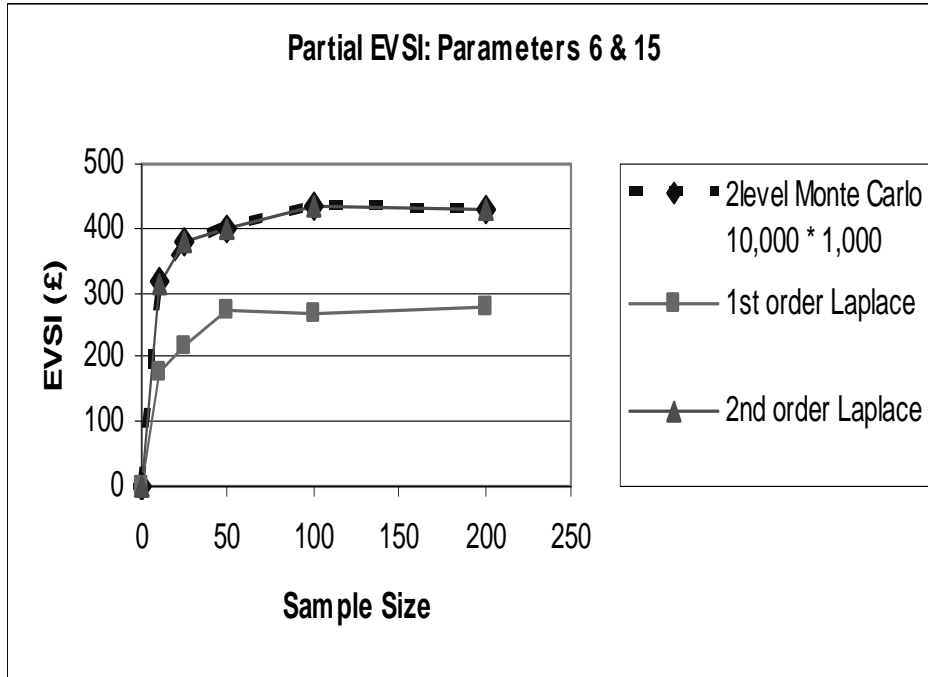


Figure 5. EVSI results for a proposed data collection exercise on parameters (θ_6, θ_{15}) using 1st order and 2nd order Bayesian Approximation compared with traditional 2 level Monte Carlo simulation

The comparison we need to make concerns the evaluation of the maximum of inner expectations over the possible treatments when using either the Bayesian approximation or the Monte Carlo sampling approach. Figure 6 shows the comparison of this maximum of inner expectation results for T0 and T1 for 1000 different simulated data-sets, again for parameters θ_6 and θ_{15} . Out of the 1000 simulated data-set samples, the resulting decision between the 2 treatments was different in just 7. When the decision was different it was in the case where the value of the net benefit of T1 given the data and T0 given the data were very close, so that the impact on the EVSI calculation itself of this 0.7% decision error is very small.

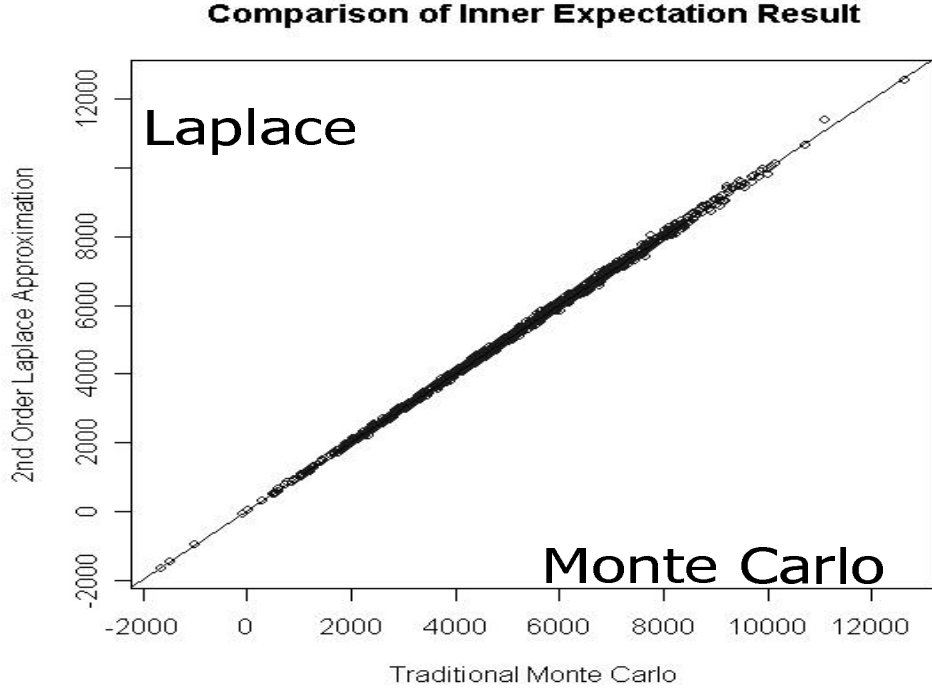


Figure 6. Comparison of the evaluated maximum of inner integral approximation for parameters (θ_6, θ_{15}) in Case Study 2, sample size $n=50$, using 2nd order Bayesian Approximation versus traditional Monte Carlo sampling on the same 1000 simulated data-sets

It is instructive to understand why the 1st term approximation is inaccurate in case study 2. The posterior probability distribution remains symmetric in any single dimension of θ , and the α_{\pm} are $= \frac{1}{2}$. Thus the second term is made up of d versions of computations of the form $\frac{1}{2}NB(t, \theta_i^-) + \frac{1}{2}NB(t, \theta_i^+) - NB(t, \hat{\theta})$. However, because of correlations between the model parameters, the sum-product form of the net-benefit functions now produces non-linearity. Thus, $\frac{1}{2}NB(t, \theta_i^-) + \frac{1}{2}NB(t, \theta_i^+) \neq NB(t, \hat{\theta})$. For example, consider that θ_5 and θ_7 are positively correlated and the net benefit function is $NB(T0) = \lambda(\theta_5\theta_6\theta_7 + \theta_8\theta_9\theta_{10}) - (\theta_1 + \theta_2\theta_3\theta_4)$. Now if we wish to evaluate the net benefit function at θ_5^+ which might have a value for θ_5 which is say double $\hat{\theta}_5$ i.e. $2\hat{\theta}_5$, then because θ_7 is also positively correlated, the associated θ_7 will be higher than $\hat{\theta}_7$ and the net benefit evaluated will more than double.

3.3 Computational Efficiency Gains

The computation time savings in our relatively simple case studies are smaller than will be achieved in models with more complex net benefit functions or non-conjugate probability distributions. Measured using a Pentium 4 1.8GHz personal computer with 512Mb RAM, the computation time to evaluate a 2level estimate of EVSI with 10,000 outer and 10,000 inner samples was 39.9 hours. This was made up of 2 main elements. Firstly, the time to produce one outer parameter sample, its associated data-set sample and subsequent Bayesian updating of multi-variate normal parameters analytically was 0.00858 seconds (around 120 of these per second). Secondly, the time to compute one inner integral Monte Carlo sample for both treatments together was 0.00144 seconds (around 700 of these per second), resulting in a time to produce an estimate of the inner expectation based on 10,000 samples of 14.4 seconds.

In contrast the Laplace approximation method takes more time for the outer sampling and associated calculations but significantly less time to estimate the inner expectation. The time to produce one outer parameter sample, its associated data-set sample and the computation of the θ^\pm matrices and α^\pm vectors of weights was 3.030 seconds. This must be done for each simulated data-set. For the inner integral though, significant savings are made because the evaluation of both net benefit functions is only undertaken $3d+1$ i.e. 58 times, rather than the 10,000 occasions for our Monte Carlo evaluation. This take 0.08 seconds as opposed to 14.4. Thus having calculated the θ^\pm matrices and α^\pm vectors and evaluated the inner expectation in 3.11 seconds combined the Laplace approximation approach in our case study was overall 4.6 times faster than the traditional method.

Efficiency gains in more complex models are likely to be higher than in our case studies. Table 3 shows that if the complexity of the model increases so that the net benefit functions took $\frac{1}{100}$ th, $\frac{1}{10}$ th, or over 1 second then the efficiency gain improves up to around 170 times faster. Similarly, if the Bayesian updating process required say 1 minute or even one hour using MCMC, then the efficiency could be up to 1,000 times faster.

Efficiency Gain What-If Analysis							
		Time for outer sample and Bayes update (seconds)					
		0.00858	0.01	0.1	1	60	3600
Time to evaluate net benefit function (seconds)	0.00144	4.6	4.6	4.6	4.9	24	1,161
	0.01	28	28	28	28	44	1,025
	0.1	113	113	113	113	120	521
	1	164	164	164	164	165	223
	60	172	172	172	172	172	173
	3600	172	172	172	172	172	172

Table 3. EVSI Computation Time Efficiency Gain (4.6 means Bayesian approximation is 4.6 times faster)

4 Discussion

In this paper we have developed some new formulae for posterior expectations based on the asymptotic theory of signed root log-density ratios. These approximations are an important adaptation of Sweeting and Kharroubi (2003) designed to absorb the prior into the likelihood function, constructing formulae that focus on the posterior. One immediate practical value of this novel approximation lies in reducing the number of computations and the time required when applied to the calculation of the expected value of sample information from decision theory. The case study models demonstrate the feasibility of the new approach, showing that expected value of sample information approximated results are similar to the standard nested Monte Carlo sampling method, but are achieved with significant computation time reductions.

The approximation formula consists of a first order term and a second, more complex, correction term. The new approximations provide the possibility for a 1st term approximation to be accurate. The second term is likely to be non-negligible except in those simple situations where, the net benefit functions are linear, the model parameters are independent and the probability distributions for the model parameters are symmetric. In practice therefore either of the Laplace approximation approaches, that based

on SRLDR presented here or the one based on SRLLR (Sweeting and Kharroubi 2003) would be applicable. In computation terms exactly the same types of processes need to be undertaken and efficiency gains between the two are probably not significant. Both approaches have the advantage that they are independent of the function $v(\theta)$, which contrasts with earlier Laplace approximation formulae from Tierney and Kadane (1986), which required re-computation if the function $v(\theta)$ changed in some way. We prefer the new approach to the extent that it provides a more intuitive 1st term i.e. $v(\hat{\theta}_{posterior})$.

The approximation approach is valid for any form of net benefit function as long as we have a finite number of decision strategy options to compare. It works for analytic models i.e. health economic evaluation models with a relatively complex system of assumptions and model parameters, effectively producing a large analytic formula for net benefit, say through a spreadsheet structure. The approximation also works for stochastic models. For example, an individual level simulation model, which produces estimates of expected net benefit for the population by sampling say 10,000 individual patients receiving the different treatments. Such models are often much more computationally intensive than analytic models but the Laplace approximation method still applies and indeed has much greater computational savings since the computationally expensive net benefit functions are computed many fewer times. The decision model case studies presented in this paper are of a very simple sum-product net benefit functional form. The posterior mode was calculated analytically, using the conjugate nature of the multi-variate normal distribution, and only the first term of formula (6) was needed to provide adequate estimates in a much shorter time for case study 1, when all parameters were independent.

The approximation approach is also valid for any probability distributions used to characterise parameter uncertainty as long as they are smooth, differentiable functions. This will cover the vast majority of characterised uncertainty in health economic models but does exclude discrete distributions including those in the form of a non-parametric histogram. The essential requirement is that we can determine the various components in formula (6) via numerical optimisation techniques. Thus, we need to be able to differentiate the log posterior density i.e. get $\ell'(\theta)$, $\ell''(\theta)$ and invert the information matrix J . One might ask whether the approximation could also apply to the outer expectation. Unfortunately, because the inner function of θ is now expressed in terms of $\hat{\theta}$, θ_i^+ and θ_i^- which are solutions of equations, we cannot write down the expectation in an analytic function of the form $E(v(\theta)|X)$ and hence the approximation formula cannot be used.

The main current limitation of the approximation theory is that it assumes that the same amount of data (i.e. n samples) will be collected on each of the parameters of interest θ_I . This is not too limiting for many real case studies e.g. how valuable would a randomised controlled trial of $n = 50$ be compared to one of $n = 300$. However it can be limiting for more complex proposed data collection exercises e.g. collect a subset of data on utilities for $n = 100$ alongside a trial of $n = 300$. We call this the 'unbalanced' case because there is a different sample size for different parameters (e.g. collecting n_1 samples on some parameters and n_2 on some others) producing a data matrix which we can consider either as no longer square, or as having missing data. This causes an immediate problem in defining the likelihood function of $X|\theta$ (and hence computing θ^\pm and α^\pm), a theoretical problem that is not addressed in the statistical literature as far as we are aware. Further developments in the theory are needed here. The solution might be some form of nested, iterative application of the approximation method (e.g. $E(v(\theta)|X_i, X_j, X_k)$). Future application of the approximation approach should also focus on more complex case studies in particular when MCMC methods are necessary under the traditional approach and when the net benefit functions are much more computationally expensive than the simple forms used here. It may also be useful to establish empirically under what circumstances the first term approximation $v(\hat{\theta})$ is accurate enough. A final methodological question is whether the approach might be used to estimate the inner integral in partial EVPI computation. Effectively the inner integral in partial EVPI is dominated by an infinite sample size data collection exercise and ignores the prior and Bayesian updating. Nevertheless, the construction of θ^\pm and α^\pm is a function of the characterised uncertainty and should still be possible. A set of [R] programs to implement the approach on a generic set of EVSI problems is available on our website. (<http://www.shef.ac.uk/chebs/>).

A series of other basic ways to increase the efficiency of EVSI computations have also evolved in our research and are worth setting out. First, it is sensible to reduce noise in the calculations as far as possible. Thus the same Monte carlo sample for the parameters of interest in the outer loop can be used for a series of increasing sample sizes to produce the simulated data. Consider the case where we wish to evaluate EVSI for $n = 10, 50, 100$ and 200. If we produce an outer sampled parameter of interest $\theta_{outersample1}$, we then can simulate a sample data-set for $n = 10$ patients $X_{10}|\theta_{outersample1}$, and then a further 40 patients so that $X_{50}|\theta_{outersample1} = X_{40}|\theta_{outersample1} + X_{10}|\theta_{outersample1}$, a further 50 again to give us 100 sampled patients, and finally a further 100 to give us $n = 200$.

Second, we have found in our case studies that the EVSI curve itself is a smooth function of sample size n , starting at zero for sample size $n = 0$, and tending upwards to (and bounded above by) the partial EVPI for the parameters of interest. This curve, at least in our case studies, can be estimated by a functional form that involves the exponential of the square root of n i.e. $EVSI(n, \theta_I) = (1 - e^{-\beta\sqrt{n}}) \times PartialEVPI(\theta_I)$. Once a small number of points have been estimated it may be useful to use regression of $\ln(EVSI(n))$ against \sqrt{n} to estimate further EVSI points from the fitted curve itself rather than using either 2 level Monte Carlo sampling or our approximation approach.

The question arises - how many inner samples are needed for accurate estimation of the $[\max_t E_{(\theta_I^c, \theta_I | X_{\theta_I})} NB(t, \theta_I, \theta_I^c)]$? The answer depends firstly upon the variances of the net benefit functions and secondly on the closeness to each other of the expected net benefits of the different treatments t . For each t , we can estimate the variance in expected net benefit $E_{(\theta_I^c, \theta_I | X_{\theta_I})} NB(t, \theta_I, \theta_I^c)$ by taking a particular simulated dataset and estimating the variance produced when we estimate the inner integral for each treatment t using just one sample of conditional parameters $(\theta_I, \theta_I^c | X_{\theta_{sample}})$, calling it say σ_{ENBt}^2 . By the central limit theorem, increasing the number of inner Monte Carlo samples $jdraw$ reduces our uncertainty in the mean expected net benefit of t by a factor of \sqrt{jdraw} . That is our uncertainty in the mean of the net benefit of treatment t is measured by $\sigma_{ENBt}^2 / \sqrt{jdraw}$. Thus increasing the number of Monte Carlo samples reduces the uncertainty in our estimates of the inner expectation in proportion to \sqrt{jdraw} .

In practice, the next question is - how accurate does the estimated EVSI answer need to be? The answer depends upon the use to which the results are put. One can use EVSI simply as a refinement of partial EVPI in the context of sensitivity analysis measurement. Partial EVPI allows the analyst to quantify how much of the decision uncertainty is a consequence of our current uncertainty around particular parameters. For example, the uncertainty in θ_6 and θ_{15} represents around 27% of overall EVPI in our case study. Using partial EVSI we can see that a study of size $n = 50$ has a value around 25% of overall EVPI. This helps to provide a 'feel' both for how important current parameter uncertainty is and how amenable to reduction via realistically sized data collection exercises the uncertainty might be. However, the concept of valuing reductions in uncertainty in EVSI is not just relative but absolute. The aim is sometimes to use EVSI calculations to perform quantified trade-offs against costs of research. Here, the analyst builds a second model, a model of the costs of proposed data collection,

which might be as simple as a fixed cost of undertaking a study plus a marginal cost related to the sample size n . The value of the data collection per person affected by the decision choice is given by the EVSI calculations described in this paper. This must then be multiplied by the 'decision prevalence' i.e. the numbers of people affected over the 'lifetime' of the decision choice. The optimal n is defined mathematically at the point where the societal value of the data collection exercise minus the cost of data collection reaches its maximum. The level of accuracy required for the per person $EVSI(\theta_I, n)$ is therefore part of a kind of meta uncertainty problem, whereby we need also to have a good understanding of the costs of data collection and decision prevalence numbers.

These trade-offs raise four wider methodological issues for the implementation of EVSI approaches. Firstly, decision makers are often keen to establish not just the optimal n for a particular study, but the best value from a range of research study options with limited research funds. They might well be minded to consider the cost-effectiveness ratios of research proposals i.e. expected cost divided by expected societal value taking into account too the timing of the data collection and its effect on sequential decisions. Secondly, there is no single global decision maker considering the adoption of health technologies, but the value of any research undertaken may well have global rather than national impact, meaning that application of the approach in particular jurisdictions may be underestimating the value obtained globally. Thirdly, in the framework presented here, we assume it is a societal decision maker funding the research and trading off its costs and benefits. Since much data collection is undertaken by pharmaceutical or medical technology developers in the commercial sector, we can conceptualise a direct commercial net benefit as opposed to societal net benefit function, which would demand a model of expected sales or profit accruing following different simulated data-set results. Finally, alongside these issues there remains the fundamental issue that inadequate modelling or inadequate characterisation of uncertainty through our ignorance of unknown factors provides a limit on uncertainty analysis in general.

In conclusion, we have developed a novel Bayesian approximation formula for posterior expectations of real valued functions given observed data $E(v(\theta)|X)$. This powerful method has been applied to the context of EVSI in decision theory. It is clear that EVSI calculations using our new Laplace approximation are able to accurately replace the traditional 2 level Monte Carlo sampling algorithm computation in the case studies examined. Our case studies focus on proposed data collection exercises with the same number of samples on every parameter and further development would be useful for the

rarer 'unbalanced' case. The approximation method is very generalisable, working for any net benefit function including stochastic simulation models and any smooth mathematically defined probability distribution. Computation time reductions are significant, with greater savings for more complex, computationally expensive decision models and those requiring MCMC for the Bayesian updating. We hope that health economists, modellers and statisticians will be able to take the method forward and to apply EVSI in a much wider range of decision model problems than has hitherto been possible.

Appendix 1: Derivation of Novel Bayesian Approximation Formula (6)

Here we briefly derive formula (6) for posterior expectations based on the signed root log-density ratio. The reader may refer back to Sweeting and Kharroubi (2003) for more details. Define

$$z^i = z^i(\theta_i) = \{k^i(\theta_{i-1})\}^{1/2} \{\theta^i - \hat{\theta}^i(\theta_{i-1})\} \quad (9)$$

where $k^i(\theta_i) = -\partial^2 l(\theta_i) / (\partial \theta^i)^2$. Let

$$Q^i(z_i) = \{-z^i / l_i(\theta_i)\} \{\nu_i(\theta_i) / \nu_{i-1}(\theta_{i-1})\}.$$

where $l_i(\theta) = \partial l(\theta) / \partial \theta^i$, $\nu_i(\theta) = |j^{(i+1)}(\theta)|^{-1/2}$ and $j^{(i)}$ is the submatrix of j corresponding to $\theta^{(i)}$. Now let θ^{i-} , θ^{i+} be the solutions to the equations $z^i(\hat{\theta}_{i-1}, \theta^i) = -1$, $z^i(\hat{\theta}_{i-1}, \theta^i) = 1$ respectively; that is $\theta^{i-} = \hat{\theta}_{i-1} - (k^i)^{-1/2}$ and $\theta^{i+} = \hat{\theta}_{i-1} + (k^i)^{-1/2}$, where k^i is the reciprocal of the first entry in $\{J^{(i)}\}^{-1}$. Write $\theta_i^- = (\hat{\theta}_{i-1}, \theta^{i-})$ and $\theta_i^+ = (\hat{\theta}_{i-1}, \theta^{i+})$. It follows from Sweeting and Kharroubi (2003) that the computation of Bayesian Bartlett corrections can be achieved as follows

$$s^i = \frac{\frac{1}{2}\{Q^i(-e_i) + Q^i(e_i)\}}{\frac{1}{2}\{-r^i(\theta_i^-)\}^{-1} + \{r^i(\theta_i^+)\}^{-1}} = |J^{(i)}|^{1/2} \tau^i \{\omega^i\}^{-1},$$

where e_i is the i -dimensional vector $(0, \dots, 0, 1)$, $\tau^i = \{\nu_i(\theta_i^-) / l_i(\theta_i^-)\} + \{-\nu_i(\theta_i^+) / l_i(\theta_i^+)\}$ and $\omega^i = \{-r^i(\theta_i^-)\}^{-1} + \{r^i(\theta_i^+)\}^{-1}$. Also Sweeting and Kharroubi (2003) obtain the approximation

$$\int L(\theta) \lambda(\theta) d\theta = (2\pi)^{d/2} |J|^{-1/2} L(\hat{\theta}) \lambda(\hat{\theta}) \prod_{i=1}^d s^i$$

to the normalizing constant of the posterior density. This can be used to compute an approximation to the posterior expectation of a general function $v(\theta)$. This leads to the formula

$$E\{v(\theta) | X\} \doteq v(\hat{\theta}) \prod_{i=1}^d \frac{s^{*i}}{s^i}, \quad (10)$$

where $s^{*i} = |J^{(i)}|^{1/2} \{\omega^i v(\hat{\theta})\}^{-1} \tau^{*i}$ and $\tau^{*i} = \{\nu_i(\theta_i^-) v(\theta_i^-) / l_i(\theta_i^-)\} + \{-\nu_i(\theta_i^+) v(\theta_i^+) / l_i(\theta_i^+)\}$. Let $\alpha_i^- = (\tau^i)^{-1} \{\nu_i(\theta_i^-) / l_i(\theta_i^-)\}$ and $\alpha_i^+ = (\tau^i)^{-1} \{-\nu_i(\theta_i^+) / l_i(\theta_i^+)\}$. Then, from (10), we see that

$$E\{v(\theta)|X\} \doteq \hat{v} \prod_{i=1}^d \left\{ \frac{\alpha_i^- v_i^- + \alpha_i^+ v_i^+}{\hat{v}} \right\},$$

where $v_i^- = v(\theta_i^-)$, $v_i^+ = v(\theta_i^+)$, $\hat{v} = v(\hat{\theta})$. We can also deduce the alternative summation form

$$E\{v(\theta)|X\} \doteq \hat{v} + \sum_{i=1}^d (\alpha_i^- v_i^- + \alpha_i^+ v_i^+ - \hat{v}).$$

From this, the results exists.

References

Ades, A. E., Lu, G. and Claxton, K. (2004). Expected Value of Sample Information Calculations in Medical Decision Modeling. *Journal of Medical Decision Making* **24**, 207–227.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.

Brennan, A., Kharroubi, S. A., Chilcott, J. and O’Hagan A. O. (2002a). A two level monte carlo approach to calculation expected value of sample information: how to value a research design. Presented at the 24th Annual Meeting of the Society for Medical Decision Making, October 23rd 2002, Washington.

Brennan, A., Chilcott, J., Kharroubi, S. A. and O’Hagan, A. (2002b) A two level Monte Carlo approach to calculating expected value of sample information:- How to value a research design. Discussion paper, 2002. . Available at <http://www.shef.ac.uk/content/1/c6/02/96/29/>

Brennan, A., Kharroubi, S. A., Chilcott, J. and O’Hagan A. O. (2002c). A two level Monte Carlo approach to calculating expected value of perfect information:- Resolution of the uncertainty in methods. Discussion Paper; <http://www.shef.ac.uk/content/1/c6/02/96/05/Brennan.>

Briggs, A. H., and Gray, A. (1999). Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technology Assessment*, **3**.

Chilcott, J., Brennan, A., Booth, A., Karnon, J. and Tappenden P. (2003). The role of modelling in prioritising and planning clinical trials. *Health Technology Assessment*, **7**(23).

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

- Claxton, K. (1999). The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*, **18**, 341–364.
- Claxton, K., Lacey, L. F. and Walker, S. G. (2000). Selecting treatments: a decision theoretic approach. *Journal of the Royal Statistical Society A* **163**, 211–226.
- Claxton K, and Thompson K. (2001) A dynamic programming approach to efficient clinical trial design. *Journal of Health Economics* **20**, 797-822
- Coyale D, Buxton MJ, O'Brien BJ. (2003) Measures of importance for economic analysis based on decision modeling. *Journal of Clinical Epidemiology* :**56**;989-997.
- DiCiccio, T. J. and Stern, S. E. (1993). On Bartlett adjustments for approximate Bayesian inference. *Biometrika* **80**, 731–740.
- Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72**, 45–58.
- Felli, J.C., Hazen, G.B., 1998. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making* **18**, 95109.
- Kass, R. E., Tierney, L. and Kadane, J. (1989). Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika* **76**, 663–674.
- Lawley, D. N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika* **43**, 295–303.
- Lewis G, Rippon I, de Francisco A and Lipworth S. (2004) Outputs And Expenditures On Health Research In Eight Disease Areas, 1996-2001. *Global Forum for Health Research Forum 8, Mexico City, November 2004*
- Meltzer D. (2001) Addressing uncertainty in medical cost-effectiveness analysis-implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *J Health Econ.* **20**:109-129.
- Raiffa, H. and Schlaiffer, R. (1967). *Applied Statistical Decision Theory*. New York: Wiley Interscience.
- Stinnett, A. and Mullahy, J. (1998). Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analyses. *Journal of Medical Decision Making* **18**, S68–S80.
- Sweeting, T. J. (1995). A framework for Bayesian and likelihood approximations in statistics. *Biometrika* **82**, 1–23.
- Sweeting, T. J. (1996). Approximate Bayesian computation based on signed roots

of log-density ratios (with discussion). In *Bayesian Statistics 5*, J. M. Bernardo et al. (eds.), 427–444. Oxford University Press; Oxford.

Sweeting, T. J. and Kharroubi, S. A. (2003). Some new formulae for posterior expectations and Bartlett corrections. *Test* **12**, 497–521.

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

Tierney, L. & Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.*, **81**, 82-86.