Calculating the Expected Value of Sample Information in Practice: Considerations from Three Case Studies

Anna Heath, Natalia R. Kunst, Christopher Jackson, Mark Strong, Fernando Alarid-Escudero, Jeremy D. Goldhaber-Fiebert, Gianluca Baio, Nicolas A. Menzies, Hawre Jalal, and on behalf of the Collaborative Network for Value of Information (ConVOI)

Abstract

Investing efficiently in future research to improve policy decisions is an important goal. Expected Value of Sample Information (EVSI) can be used to select the specific design and sample size of a proposed study by assessing the benefit of a range of different studies. Estimating EVSI with the standard nested Monte Carlo algorithm has a notoriously high computational burden, especially when using a complex decision model or when optimizing over study sample sizes and designs. Therefore, a number of more efficient EVSI approximation methods have been developed. However, these approximation methods have not been compared and therefore their relative advantages and disadvantages are not clear.

A consortium of EVSI researchers, including the developers of several approximation methods, compared four EVSI methods using three previously published health economic models. The examples were chosen to represent a range of real-world contexts, including situations with multiple study outcomes, missing data, and data from an observational rather than a randomized study. The computational speed and accuracy of each method were compared, and the relative advantages and implementation challenges of the methods were highlighted.

In each example, the approximation methods took minutes or hours to achieve reasonably accurate EVSI estimates, whereas the traditional Monte Carlo method took weeks. Specific methods are particularly suited to problems where we wish to compare multiple proposed sample sizes, when the proposed sample size is large, or when the health economic model is computationally expensive. All the evaluated methods gave estimates similar to those given by traditional Monte Carlo, suggesting that EVSI can now be efficiently computed with confidence in realistic examples.

Introduction

The Expected Value of Sample Information (EVSI) quantifies the expected benefit of undertaking a potential future study that aims to reduce uncertainty about the parameters of a health economic model. The expected net benefit of sampling (ENBS), which is the difference between EVSI and the
expected research study costs, can be used to inform decisions regarding study design and research prioritization. The future study with the highest ENBS should be prioritized if we wish to maximize economic efficiency. Thus, EVSI has the potential to determine the value of future research and to guide its design when accounting for economic constraints.

Despite this potential, EVSI has rarely been used in practical settings for a variety of reasons \cite{3}. In the past, calculating EVSI in real-world scenarios has been based on nested Monte Carlo (MC) sampling \cite{4}, and this is computationally costly if we wish to produce accurate estimates with high precision. This computational burden is further increased when one aims to compute EVSI for multiple trial designs in order to determine the optimal (i.e., with the highest ENBS) research study \cite{5,6}. High performance computing resources can be used to overcome some of these barriers, but often at the expense of an increased requirement for programming skills and an increase in the complexity of the analysis.

Several methods have been developed to overcome these computational barriers and unlock the potential of EVSI as a tool for research prioritization and trial design optimization \cite{7,8,9,10,11,12,13,14,15,16,17}. However, as many of these methods have been developed concurrently, they have not been compared. Additionally, EVSI estimation methods are typically evaluated using health economic models and trial designs chosen for computational convenience rather than those that reflect real-world decision making.

Some of the EVSI estimation methods that have been proposed place restrictions on the structure of the underlying health economic model and/or the study design \cite{7,8,9}. These restrictions typically take the form of an assumption about the study data that ensures that the prior and posterior model parameter distributions take the same form (conjugacy), and by doing so, allow for computationally efficient EVSI estimation. This, however restricts the applicability of these methods. EVSI estimation based on minimal modelling, where a comprehensive clinical trial is available to inform EVSI estimation, has also been proposed \cite{18}. However, this paper aims to review EVSI estimation procedures for three case studies where the health economic models are based on a diverse evidence base and, when combined with the proposed study designs, do not fulfill the assumptions required for these restrictive or minimal modelling methods.

Thus, our comparison is restricted to four recent calculation methods developed by (in chronological order) Strong et al. \cite{10}, Menzies \cite{11}, Jalal and Alarid-Escudero \cite{13} (extending a method proposed in Jalal et al. \cite{12}), and Heath et al. \cite{15,16,17}. Whilst these methods are all based on different approaches and assumptions, they all provide estimation techniques for approximating EVSI that, in comparison to nested MC sampling methods, are less computationally demanding whilst retaining accuracy.

Our primary goal is to test the four EVSI estimation methods across a range of health economic models and trial designs to gain a greater understanding of their behaviour in practice. We will evaluate the accuracy of EVSI estimation methods across the three models and the computational time required to obtain these estimates. These three models have several key features that reflect real-world trial design and may make it challenging to estimate EVSI in practice. These are: the presence of multiple trial outcomes, missingness or loss to follow-up in the data, and a study design that is observational rather than randomized.

**Notation and Key Concepts**

Health economic decision making aims to determine the intervention, from some set of feasible alternatives, that is expected to be optimal in terms of utility (which is usually net monetary benefit or net health benefit \cite{12}). We characterize a health economic model as a function that takes as input a vector of parameters $\theta$, and returns the costs and health effects associated with each intervention in the set of alternatives. Uncertainty in the input parameters is represented using a probability distribution $p(\theta)$. To find the optimal intervention, costs and effects are combined into a single measure of economic value by calculating the net benefit for each of the $T$ treatment options considered relevant, conditional on $\theta$. Uncertainty about $\theta$ induces uncertainty about the net benefit for each treatment $t = 1, \ldots, T$. We denote the net benefit for treatment $t$ given parameters $\theta$ as $\text{NB}_t^\theta$. Under the assumption of a rational, risk neutral decision maker, the optimal intervention given current evidence is the intervention associated with the maximum expected net benefit.

We consider that the model parameters can be split into two sets $\theta = (\phi, \psi)$, where $\phi$ is a subset of parameters that we wish to obtain more information on, and $\psi$ are the remaining parameters. For example, clinical trials are informative for clinical outcomes but may not collect information about health state utilities or costs. The economic value of eliminating all uncertainty about $\phi$ (assuming risk neutrality) is equal to the Expected Value of Partial Perfect Information (EVPPPI) \cite{20,21,22}. This is
given by

\[
\text{EVPPI} = E_{\phi} \left[ \max_t \mathbb{E}_{\theta|X} \left[ \text{NB}_{t}^{\theta} \right] \right] - \max_t \mathbb{E}_{\theta} \left[ \text{NB}_{t}^{\theta} \right].
\]  

(1)

The EVSI is the value of collecting additional data, denoted \( X \), to inform the parameters \( \phi \) and is bounded above by the EVPPI. If these data had been collected and observed to have a value \( x \), they would be combined with the current evidence to generate an updated distribution for \( \phi \), \( p(\phi \mid x) \). Under a Bayesian approach, this would in turn be used to update the distribution of the net benefit of each treatment. The optimal intervention conditional on the data \( x \) is the treatment associated with the maximum expected net benefit based on the updated knowledge about the relevant parameters \( \phi \). If the optimal intervention changes, compared to the current decision, then the information in \( x \) has value. However, as the data have not been collected yet (and may never be), the average value over all possible datasets is considered. Mathematically, EVSI is defined as

\[
\text{EVSI} = E_X \left[ \max_t \mathbb{E}_{\theta|X} \left[ \text{NB}_{t}^{\theta} \right] \right] - \max_t \mathbb{E}_{\theta} \left[ \text{NB}_{t}^{\theta} \right],
\]

(2)

where the distribution of \( X \) can be defined through \( p(X, \theta) = p(\theta)p(X \mid \theta) \) where \( p(X \mid \theta) = p(X \mid \phi) \) is the sampling distribution for the data given the parameters. We assume that the sampling distribution for the data is only defined conditional on \( \phi \), i.e., does not provide information on the value of the parameters \( \psi \), except through any relationship with \( \phi \).

**Calculation Methods for EVSI**

It is rarely possible to compute EVSI analytically as the net benefit is often a complex function of \( \theta \). Additionally, it is challenging to compute the expectation of a maximum analytically as required in the first term of equation (2). Therefore, a range of methods have been developed to approximate EVSI.

**Nested Monte Carlo Computations for EVSI**

The simplest approximation method [4] computes all the expectations in equation (2) using MC simulation. The second term can be computed by simulating \( s = 1, \ldots, S \) parameter values, \( \theta_s \), from \( p(\theta) \). The simulated values are used as inputs to a health economic model to obtain \( S \) simulations of the net benefit for each intervention, denoted \( \text{NB}_{t}^{\theta_s} \). Note that this process is required to perform a “probabilistic sensitivity analysis” (PSA) [23], used to assess the impact of parametric uncertainty on the decision uncertainty, which is mandatory in various jurisdictions [24, 25, 26]. The average of \( \text{NB}_{t}^{\theta_1}, \ldots, \text{NB}_{t}^{\theta_S} \) for each intervention can be computed and \( \max_t \mathbb{E}_{\theta} \left[ \text{NB}_{t}^{\theta} \right] \) is estimated by the maximum of these means.

The first term in equation (2) is more complex to compute by simulation. Firstly, \( S \) datasets \( X_s \) must be generated conditional on the simulated \( \theta_s \) from the assumed sampling distribution \( p(X \mid \theta_s) \). For each \( X_s \), we simulate \( R \) values from the updated distribution of the model parameters \( p(\theta \mid X_s) \). These \( R \) simulations are used as inputs to the health economic model to simulate from the updated distribution of the net benefit for each intervention. The mean net benefit for each treatment option is then calculated to estimate \( \mathbb{E}_{\theta|X} \left[ \text{NB}_{t}^{\theta} \right] \) for \( t = 1, \ldots, T \). The maximum of these simulated means is then selected for each \( X_s \). Thus, to compute EVSI by MC simulation, we require \( S \times R \) runs of the health economic model. This is computationally expensive for standard choices of \( S \) and \( R \), which are typically in the thousands. Therefore, the following methods focus on approximating the updated mean of the incremental net benefit associated with each intervention \( t \) using a smaller simulation burden. We denote the expectation of the incremental net benefit, conditional on data \( X \), as

\[
\mu_t^X = E_{\theta|X} \left[ \text{NB}_{t}^{\theta} \right].
\]

In a similar manner, we also denote the expectation of the incremental net benefit, conditional on some value of the parameters of interest \( \phi \), as

\[
\mu_t^\phi = E_{\theta|\phi} \left[ \text{NB}_{t}^{\theta} \right].
\]

Finally, to increase the numerical stability of the following approximation methods, it is easier to work in terms of the incremental net benefit or loss, defined, without loss of generality, as \( \text{INB}_{t}^{\theta} = \text{NB}_{t}^{\theta} - \text{NB}_{1}^{\theta} \) for \( t = 2, \ldots, T \).
Strong et al.

The Strong et al. method estimates EVSI by fitting a regression model between the simulated values of the incremental net benefit, as the ‘dependent’ or ‘response’ variable, and a scalar or low-dimensional summary for the simulated dataset \( X \) as the ‘independent’ or ‘predictor’ variable(s) \([16]\). This low-dimensional summary for \( X \) should reflect how the data would be summarized if the study were to go ahead and must be computed for each simulated dataset \( X_s \). Once this regression model has been fitted, \( \mu_t^X \) is estimated by the fitted values from this regression model. EVSI is then estimated directly from these estimates of \( \mu_t^X \).

Menzies

Menzies \([11]\) presents two EVSI estimation methods, the most accurate of which estimates \( \mu_t^X \) by reweighting simulations of \( \mu_t^\phi \). This reweighting is based on the likelihood of observing a simulated dataset \( X \) conditional on different values for \( \phi \). The term likelihood is used in the statistical sense and is equal to \( p(X | \phi) \).

This method simulates \( S \) future datasets \( X_s \) from \( p(X | \phi_s) \). The likelihood for every simulated vector for \( \phi \) is then calculated conditional on \( X_s \). For the sample \( X_s \), \( \mu_t^{X_s} \) is estimated as the average of \( \mu_t^\phi \), weighted by the likelihood of the dataset \( X_s \), and the method can therefore be seen as an example of importance sampling \([27, 28]\). EVSI is estimated based on the estimate of \( \mu_t^{X_s} \) for each future sample.

Jalal et al.

The Jalal et al. method published by Jalal and Alarid-Escudero \([13]\), building on work from Jalal et al. \([12]\), fits a linear meta-model \(^1\) between the simulated incremental net benefit values, as the response variable, and simulations for \( \phi \), as the predictor variables. Each term of the linear meta-model is then rescaled based on a Gaussian-Gaussian Bayesian updating approach to estimate its “posterior” expectation across different future datasets \( X \). These estimated distributions are then recombined using the coefficients of the linear model to estimate \( \mu_t^X \) and compute EVSI.

For a proposed future data collection strategy of size \( N \), the rescaling factor for each term of the linear meta-model is equal to

\[
\frac{N}{N+N_0},
\]

where \( N_0 \) is known as the prior effective sample size. In some prior-likelihood pairs, \( N_0 \) can be obtained analytically. In other settings, \( N_0 \) can be estimated using one of two estimation methods. If the data \( X \) can be summarized using a summary statistic \( W(X) \), then \( N_0 \) can be computed as a function of the variance of \( W(X) \). If a suitable statistic cannot be derived, then nested posterior sampling can be used to estimate \( N_0 \). In this method, \( S \) future datasets \( X_s \), \( s = 1, \ldots, S \) are simulated. Each of these samples is used to update the information about the model parameters \( p(\theta | X_s) \), typically using \( R \) simulations and computing the mean for \( \phi \). The variance of the mean for \( \phi \), across different samples \( X_s \), is then used to estimate \( N_0 \). Computationally, this nested sampling method to compute \( N_0 \) is relatively computationally expensive compared to the other two proposals to determine \( N_0 \). However, calculation of \( N_0 \) is only needed once to compute EVSI across study size.

Heath et al.

The Heath et al. \([16, 17]\) estimation method combines the simulations \( \mu_t^\phi \) and a modified nested MC sampling method to estimate EVSI. This method reduces the number of times the updated distribution of the net benefit must be simulated to estimate EVSI from \( S \), typically at least 1000, to \( Q \), usually between 30 and 50 \([17]\). Thus, EVSI is estimated with \( Q \times R \) health economic model runs.

The Heath et al. method uses nested MC sampling to estimate the variance of the incremental net benefit for different future datasets. These estimated variances rescale simulations of \( \mu_t^\phi \) for \( t = 2, \ldots, T \) to approximate simulations of \( \mu_t^X \) which can be used to estimate EVSI. The Heath et al. method only requires a single nested simulation procedure to estimate EVSI across sample size \([17]\).

\(^1\)A “linear” model is required for this method. However, non-linear functions of \( \phi \) can be defined and combined linearly to account for flexible relationships between the incremental net benefit and the parameters \( \phi \).
Case Studies

These EVSI methods are applied to three case studies designed to explore trial designs using health economic models that make EVSI estimation reflective of real-world decision making. The first case study is a stylized chemotherapy example used to evaluate EVSI estimation in the presence of multiple outcomes, reflecting a realistic trial design with a single primary, and multiple secondary, outcomes. The second case study evaluates EVSI methods in the presence of missingness in the data using a previously published health economic model to explore EVSI estimation when we account for standard considerations in trial design and development. Finally, we evaluate EVSI methods for a health economic model based on a time-dependent natural history model where the main data source is observational.

Case Study 1: A New Chemotherapy Treatment

This model was developed in Heath and Baio [16] to evaluate two chemotherapy interventions, i.e., the current standard of care and a novel treatment that reduces the number of adverse events. These two options are equal in their clinical outcomes so we focus on the adverse events. The probability of adverse events for the standard of care is denoted $\pi_0$ and $\rho$ denotes the proportional reduction in the probability of adverse events with the novel treatment.

All patients incur a treatment cost of £110 for the standard of care or £420 for the novel treatment. Patients without adverse events or those that have recovered have a quality of life (QoL) measure of $q$.

The health economic impact of adverse events is modelled with a Markov model depicted in Figure 1. In this model, $\gamma_1$ and $\gamma_2$ denote the constant probability of requiring hospital care and dying, respectively, and $\lambda_1$ and $\lambda_2$ denote the constant probability of recovery given that an individual remains at home or enter hospital, respectively. The cycle length is 1 day and the time horizon is 15 days. Recovered patients incur no further cost while patients who die have a one-time cost of terminal care. There are costs and QoL measures associated with home and hospital care. PSA distributions for the model parameters are informed using previous data or defined using expert opinion with all distributional assumptions given in the supplementary material.

![Figure 1: A four state Markov model used to model the health economic impact of adverse events from a chemotherapy treatment.](image)

Sampling Distributions for $X$

The EVSI is computed for a future two-arm randomized control trial whose primary outcome is the number of adverse events. As a secondary set of measures, the study monitors the treatment pathway for patients who experience adverse events. Thus, the trial directly informs six model parameters $\phi = (\pi_0, \rho, \gamma_1, \gamma_2, \lambda_1, \lambda_2)$ by collecting six outcomes. We will enrol 150 patients per arm.

To define the sampling distribution for the six outcomes, we model the number of adverse events using binomial distributions conditional on $\pi_0$ and $\rho$;

$X_{AE_0} \sim Bin(150, \pi_0)$ and $X_{AE_1} \sim Bin(150, \rho\pi_0)$.

The number of patients treated in hospital and the number of patients who die are modelled as

$X_{Hosp} \sim Bin(X_{AE_0} + X_{AE_1}, \gamma_1)$ and $X_{Death} \sim Bin(X_{Hosp}, \gamma_2)$. 


Finally, recovery time for patients who experience adverse events but recover is modelled with an exponential distribution conditional on the transition probabilities $\lambda_1$ and $\lambda_2$,

$$T_{HC} \sim \text{Exponential}(\eta_1)$$

with $\eta_1 = -\log(\lambda_1)$ and $i = 1, \ldots, X_{AE_0} + X_{AE_1} - X_{Hosp}$. The recovery time for every patient who recovers in hospital is modelled as

$$T^j_H \sim \text{Exponential}(\eta_2)$$

with $\eta_2 = -\log(\lambda_2)$ and $j = 1, \ldots, X_{Hosp} - X_{Death}$.

**Case Study 2: A Model for Chronic Pain**

This example uses a cost-effectiveness model developed by Sullivan et al. [29], and extended in Heath et al. [17], to evaluate treatments for chronic pain. This is based on a Markov model with 10 states, where each state has an associated QoL and cost. The model is initiated when a cohort of patients receive their initial treatment for chronic pain. Patients can experience adverse events due to treatment and can withdraw from treatment due to adverse events or lack of efficacy. Following this, they can be offered an alternative therapy or withdraw from treatment. If they withdraw from this second line of treatment, they can receive further treatment or discontinue, both considered absorbing states as the model does not consider a death state.

As a treatment for chronic pain, a patient can first either be offered morphine or an innovative treatment. If they withdraw, they are offered oxycodone as an alternative treatment. Thus, the only difference between the two options is the first-line treatment where the innovative treatment is more effective, more expensive and causes fewer adverse events. A more in-depth presentation of all the model parameters is given in [29] where the parameter distributions are gamma for costs and beta for probabilities and utilities. The means of these distributions are informed by relevant studies identified following a literature review and the standard deviation is taken as 10% of the underlying mean estimate.

The per-person lifetime EVSI is calculated, assuming a discount factor of 0.03 per year over 15 years.

**Sampling Distributions for $X$**

EVSI is computed for a study that investigates the QoL weights for patients who remain on treatment without any adverse events and of patients who withdraw from the first line of treatment due to lack of efficacy. The individual level variability in these two QoL weights is modelled, for simplicity, as independent beta distributions although the assumption of independence may be invalid [30]. The population level mean QoL weight, i.e., the mean of the individual level QoL distribution, is defined as the value of those two health states in the Markov Model. The standard deviations of the individual level distributions is then set equal to 0.3, for patients who remain on treatment, and 0.31, for patients who withdraw due to lack of efficacy [31]. We compute EVSI for trials enrolling 10, 25, 50, 100 and 150 patients. We assume that only a proportion of the questionnaires are returned, leading to missingness in the data.

To generate the data, a response rate of 68.7% is assumed, consistent with the return rate observed in [32]. We generate a response indicator for each patient in the trial using a Bernoulli distribution. If this indicator is 1, then we assume the patient returned the questionnaire and therefore we have observed utility scores for both states for that patient, simulated from the beta distributions specified above, conditional on the model parameters.

**Case Study 3: A Model for Colorectal Cancer**

This example uses a health economic model developed by Alarid-Escudero et al. [33] to evaluate a screening strategy for colorectal cancer (CRC) and pre-cancerous lesions known as adenomas. This model is based on a nine-state Markov model with age-dependent transition intensities which govern the onset of adenomas (pre-cancerous growths) and the risk of all-cause mortality. The onset of adenomas is modeled using a Weibull hazard conditional on age

$$l(a) = \lambda_1 g a^{\alpha - 1}$$

This sampling distribution for the data causes some minor issues for the Gibbs sampling procedure used in the JAGS program for Bayesian updating.
where $\lambda_1$ and $g$ are the shape and scale parameters of the Weibull distribution and $a$ is the age of the patient. Model parameters are calibrated to observed literature and uncertainty in the model parameters $g$ and $\lambda_1$ reflects the uncertainty in these calibration targets.

The costs and QoL associated with each health state are used to evaluate the economic burden of CRC. The screening strategy is assumed to capture patients with adenomas and early cancer so they can be operated on before the cancer progresses and becomes clinically detected. The proposed screening strategy has a sensitivity with a mean of 0.98 and a specificity with a mean of 0.87. Some members of the general population have undiagnosed adenomas and early stage CRC at the onset of the simulation.

**Sampling Distributions for $X$**

EVSI is computed for a study that investigates the onset of adenomas in the general population to inform the shape and scale of the Weibull hazard function. A cross-section of the general population aged between 25 and 90 without any screening history will be screened for the presence of adenomas with a gold standard test with 100% sensitivity and specificity. Upon enrollment, the age of the subjects is recorded to determine the age-specific risk. EVSI is computed for trials enrolling 5, 40, 100, 200, 500, 750, 1000 and 1500 participants.

To generate prospective data, we simulate the enrolment age for participants. Demographic data from Canada in 2011, obtained from the Human Mortality Database [34], were used to generate study subjects with an age distribution representative of the general population, with ages restricted between 25 and 90 years. Conditional on their age $a$, a participant has a probability $p(a) = 1 - e^{-\lambda_1 a^g}$ of having an adenoma or CRC. The outcome for a specific subject was simulated from a Bernoulli distribution conditional on $p(a_i)$

$$X_i \sim \text{Ber}(p(a_i))$$

where $a_i$ is the age of participant $i$. We assumed that there is no missing data as participants are enrolled and undergo the test at the same clinic visit and no other data are collected.

**Analysis**

Comparing the presented EVSI estimation methods is challenging as their accuracy and computational time are dependent on choices made by the modeller and the computational efficiency of the method implementation. Table 1 outlines the simulation choices that were made for the case studies. These choices were made to achieve EVSI estimates with a reasonable level of precision, while keeping the computation time manageable. For example, smaller sample sizes were necessary for models with a greater computational cost. We compared the speed and accuracy achievable by each method, and identified their relative advantages and challenges in practice.

The prior effective sample size for the Jalal et al. method needs to be computed once to estimate the EVSI across sample size. As posterior updating is slower for larger sample sizes, it is preferable to estimate $N_0$ with a small proposed sample $X$. However, the estimation of $N_0$ also relies on a Gaussian approximation so the sample size of $X$ should be sufficiently large to assume normality. Thus, the table above (Jalal et al. future sample size) highlights the sample size of $X$ used in the nested posterior sampling to estimate $N_0$ that balances accuracy and computational speed.

For the first two case studies, we computed a standard error for the EVSI estimates by recomputing the EVSI 200 times, each time with the same PSA samples, so that this standard error reflects uncertainty arising from any simulation involved in the EVSI estimation procedure itself.

To obtain the computational time for the four recent approximation methods, computations were undertaken on a computer with an i7 Intel processor with 16 GB of RAM in R version 3.5.1. The nested MC computations were undertaken on a Linux Google Compute Engine virtual machine. The computation time give below is the total time across all cores. Code to undertake the computations in this paper is available from GitHub at [https://github.com/convoigroup/EVSI-in-practice](https://github.com/convoigroup/EVSI-in-practice).
Simulation Choices | Case Study
|-----------------|-----------------|-----------------|-----------------|
|                 | Chemotherapy    | Chronic Pain    | CRC screening   |
|                 | side effects    | (2)             | (3)             |
| Initial PSA size| 100,000         | 100,000         | 5,000           |
| Number of \( \mu_t^R \) simulations from EVPI calculation | 100,000 | 100,000 | 5,000 |
| Nested simulation outer loop size | 100,000 | 100,000 | NA |
| Nested simulation inner loop size | 100,000 | 100,000 | NA |
| Strong et al. sample size | 100,000 | 100,000 | 5,000 |
| Menzies sample size | 5,000 | 2,500 |
| Jalal et al. \( N_0 \) computation method | nested posterior sampling | nested posterior sampling | nested posterior sampling |
| Jalal et al. \( N_0 \) estimation outer loop size | 1,000 | 1,000 | 5,000 |
| Jalal et al. \( N_0 \) estimation inner loop size | 10,000 | 10,000 | 5,000 |
| Jalal et al. \( N_0 \) estimation future sample size | 30 | 40 | 40 |
| Heath et al. outer loop size | 50 | 50 | 50 |
| Heath et al. inner loop size | 10,000 | 10,000 | 5,000 |

Table 1: The simulation choices to compute EVSI for the four recent approximation methods and the nested MC method for case study 1, 2 and 3.

Results

Case Study 1: Chemotherapy Side Effects

Figure 2 displays the 95% central intervals for the four faster EVSI approximation methods, with the nested MC estimate shown as a vertical line. All the methods produce EVSI estimates that are relatively close to the EVSI estimated by nested MC sampling, which we assume is accurate given the large simulation size. The 95% central interval for the Heath et al. method is the only interval that contains the “true” value, represented by the nested MC EVSI. At the same time, the Heath et al. estimate is associated with substantial variability compared to the other methods.

![Figure 2](image)

Figure 2: The mean per-person EVSI estimates, across 200 simulated estimation procedures, for the five methods under consideration for the chemotherapy example with a future sample size of 150 and willingness-to-pay of £30,000. The 95% central intervals from these 200 simulations are shown as horizontal lines and the gold standard MC estimator is shown as a vertical line.

Implementing the Strong et al. and Jalal et al. methods involves finding a flexible regression model that fits well and is computationally feasible to estimate. As there are six parameters in this example, finding such a model was relatively challenging and required examination of residual plots.
**Case Study 2: Chronic Pain**

Figure 3 shows that the 95% central intervals for the Heath et al. and the Menzies methods contain the nested MC estimate, which we assume to be accurate given the large simulation size, for all sample sizes. However, all methods are relatively close to the nested MC estimate. The Strong et al. method produced the shortest 95% central intervals while the three alternatives are relatively comparable. Note that the Menzies estimate is based on a smaller PSA simulation size but still offers similar variability compared to the other methods.

![Figure 3](image_url)

Figure 3: The mean EVSI estimates, across 200 simulated estimation procedures, for the five methods under consideration for the chronic pain example. EVSI was calculated across 5 different sample sizes for the future trial. The 95% central intervals from these 200 simulations are shown as horizontal lines and the gold standard MC estimator is shown as a vertical line.

In this example, the summary statistic used for the Strong et al. method is the geometric mean of \( X \) and \( 1 - X \). These statistics are sufficient to estimate the model parameters of the beta distribution and were derived using the Fisher-Neymann factorization theorem [35]. Summarizing \( X \) using the arithmetic mean and variance gives incorrect EVSI estimates for this case study.

**Case Study 3: Colorectal Cancer**

Figure 4 demonstrates a broad consensus among the four recent approximation methods for the CRC screening model. Nested MC simulations are not undertaken for this case study due to the computational time required to obtain suitably accurate estimates for comparison. Thus, while we can note that the four methods give similar results, we cannot assert that these EVSI estimates are “correct.”

For a sample size of 1,500, the Menzies EVSI estimate is incorrect. This is because the likelihood tends to 0 for large sample sizes making the weighted samples difficult to approximate. Furthermore, the Menzies method slightly over-estimates the EVSI for sample sizes between 500 and 1000. This is
because we only use a subset of the PSA simulations to obtain this EVSI estimate and the EVPPI, upper limit for EVSI, estimated using this subset is slightly over-estimated, judging from the full 5,000 PSA simulations.

Figure 4: EVSI estimates for the four methods under consideration for the CRC model. EVSI is calculated for 9 different sample sizes for the future trial and is plotted across sample size. The sample size is plotted on the log scale with the sample sizes marked on the natural scale. The EVPPI, computed using the Strong et al. EVPPI computation method [36], is included as a black line on this Figure.

Computational Time

Table 2 shows the computational time for the five EVSI computation methods for each of the three case studies. For the first two case studies, all four alternatives are considerably faster than the nested MC method. For the third case study, the computational cost of the underlying CRC model meant that it was not computationally feasible to use the nested Monte Carlo method.

For the first two case studies, the Heath et al. method has the lowest computation time as the underlying health economic model is fast. The Heath et al. method also estimates EVSI across multiple sample sizes simultaneously which improves the computational time for the Chronic Pain example compared to the Strong et al., and Menzies methods. For these two examples, the computation time required to fit an accurate regression model is relatively high, increasing the computation time for the Strong et al. method. The Jalal et al. method has the highest computation time as it uses nested MC simulation to calculate \( N_0 \). However, after estimating \( N_0 \), EVSI can be re-estimated for any sample size. Thus, if EVSI was to be estimated across more sample sizes, the Jalal et al. method would offer computational
savings on the Strong et al., and Menzies methods. For the Chemotherapy example, the Menzies method has a similar computational cost to the other three methods. However, it is estimated based on a reduced simulation size; if all 100,000 PSA simulations are used, the computation time is greater than 2 hours. For the Chronic Pain example, the Menzies method is noticeably slower as the computation time for the likelihood increases when the proposed sample size of \( X \) is larger.

| Case Study        | Computational Time (mins) |
|-------------------|---------------------------|
|                   | Nested MC | Strong et al | Menzies | Jalal et al. | Heath et al. |
| 1: Chemotherapy   | 60480     | 6.45        | 4.45    | 7.47         | 1.48         |
| 2: Chronic Pain   | 223200    | 12.05       | 86      | 22.27        | 2.46         |
| 3: Colorectal Cancer | *        | 27.24      | 91      | 7.17         | 492          |

Table 2: The computational time required to produce EVSI estimates for the five methods under consideration for the three case studies presented in this review.

For the CRC screening example, the Jalal et al. method is fastest because, even though \( N_0 \) is estimated through nested MC simulation, it must only be computed once to estimate the EVSI across sample size. In contrast, for the Strong et al., method, \( X \) is summarized by finding the maximum likelihood estimates (MLE) for \( g \) and \( \lambda_1 \) that must be estimated, using relatively slow computational optimization procedures, for each sample \( X_s, s = 1, \ldots, S \) and sample size. Thus, estimating the summary statistics is slow in this case study. The Heath et al. method is more computationally expensive as the underlying probabilistic sensitivity analysis for the CRC health economic model is expensive and must be rerun \( Q \times S = 250,000 \) times to compute EVSI. The computational time of the Menzies method is similar to the previous case studies.

**Discussion**

The paper uses three case studies to assess four novel methods for approximating EVSI. These methods were developed in response to the immense computational burden required to estimate EVSI using nested MC simulations. As these methods were developed concurrently, no head-to-head comparison has been undertaken. Additionally, these methods have typically been assessed using health economic models designed for computational simplicity rather than reflecting real-life decision making.

Thus, we compared these four methods using case studies designed to cover a number of different trial designs, interventions and health economic model structures that may make the EVSI estimation more challenging. In general, the EVSI estimates were accurate when the underlying assumptions for the respective methods were met, highlighting the importance of checking these assumptions. The computational complexity of these methods varies for different health economic models, different sampling distributions for the future data, and depending on whether optimization over different sample sizes is required.

In general, we find that the four methods are comparable in terms of accuracy and computational time in these more realistic situations. However, it should be noted that appropriately assessing accuracy is challenging because differences in the EVSI estimate could lead to alternative future research recommendations, even when the difference is small. This is especially true for diseases with high incidence as the EVSI is multiplied by the incidence to determine whether the trial offers value for research investment. The determination of whether the EVSI is sufficiently precise will depend on the decision problem at hand, so care should be taken when interpreting these results.

It is likely to be more useful to compare these methods on their ease of implementation. The “optimal” estimation method trading off accuracy, precision, computational time and ease of implementation will change depending on the health economic model structure, proposed trial design and analyst expertise. Due to the differences between these four methods and the inherent differences in health economic models and trial designs, giving general purpose recommendations is not simple and would not be unconditional.

Nonetheless, this analysis highlights that the Strong et al. is accurate and efficient, provided the analyst can correctly summarize the trial data and fit a regression model. The Menzies method is accurate but computationally relatively expensive for large PSA simulation sizes. The Jalal et al. method is efficient when estimating EVSI across sample size but may require nested posterior sampling when considering realistic data collection exercises. Finally, the Heath et al. method is accurate and efficient when the health economic model has a low computation time but becomes more unfeasible as the model
becomes more expensive. The Jalal et al. and Heath et al. methods required expertise in Bayesian methods for all the examples in this paper.

While further research is required to give comprehensive guidance on the situations in which each of these methods is most useful, we can conclude that, provided the underlying assumptions of the method are met, any of the four methods chosen is likely to produce reasonable estimates in reasonable amount of time.

Contributions

AH conceived the study, performed the analysis and drafted the paper; NRK advised on the study design and contributed to the analysis, results interpretation and drafting the paper; CJ advised on study design and contributed to drafting the paper; MS advised on the study design, verified the implementation of the Strong et al. method, and contributed to drafting the paper; FA-E advised on the study design, contributed to drafting the paper and verified the implementation of the Jalal et al. method; JDG-F advised on study design and contributed to results interpretation and drafting the paper; GB contributed to drafting the paper and the results interpretation; NM advised on the study design and contributed to drafting the paper; HJ conceived the study; contributed to drafting the paper and verified the implementation of the Jalal et al. method. All authors approved the final draft.

Acknowledgements

AH was funded by the Canadian Institute of Health Research through the SPOR Innovative Clinical Trial Multi-Year Grant. NRK was funded by the Research Council of Norway (276146) and LINK Medical Research. CJ was funded by the UK Medical Research Council programme MRC_MC_UU_00002/11. This paper draws on work that MS conducted while supported by a NIHR Post-Doctoral Fellowship (PDF-2012-05-258) from 2013 to 2016. FA-E was funded by the National Cancer Institute (U01- CA-199335) as part of the Cancer Intervention and Surveillance Modeling Network (CISNET). JDG-F was funded in part by a grant from Stanford’s Precision Health and Integrated Diagnostics Center (PHIND). GB was partially funded by a research grant sponsored by Mapi/ICON at University College London. NM was supported by National Institutes of Health (NIH) [R01AI112438-02]. HJ was funded by NIH/NCATS grant 1KL2TR0001856. The funding agreement ensured the authors’ independence in designing the study, interpreting the data, writing, and publishing the report. The authors would also like to thank Alan Brennan, Michael Fairley, David Glynn, Howard Thom and Ed Wilson for their comments and discussion as part of the ConVOI group.

References

[1] R. Schlaifer. *Probability and statistics for business decisions*. McGraw-Hill, 1959.

[2] H. Raiffa and H. Schlaifer. *Applied Statistical Decision Theory*. Harvard University Press, Boston, MA, 1961.

[3] L. Steuten, G. van de Wetering, K. Groothuis-Oudshoorn, and V. Retèl. A systematic and critical review of the evolving methods and applications of value of information in academia and practice. *PharmacoEconomics*, 31(1):25–48, 2013.

[4] A. Brennan, S. Kharrouri, A. O’Hagan, and J. Chilcott. Calculating Partial Expected Value of Perfect Information via Monte Carlo Sampling Algorithms. *Medical Decision Making*, 27:448–470, 2007.

[5] S. Conti and K. Claxton. Dimensions of design space: a decision-theoretic approach to optimal research design. *Medical Decision Making*, 29(6):643–660, 2009.

[6] E. Jutkowitz, F. Alarid-Escudero, K. Kuntz, and H. Jalal. The Curve of Optimal Sample Size (COSS): A Graphical Representation of the Optimal Sample Size from a Value of Information Analysis. *PharmacoEconomics*, 2019.

[7] A. Ades, G. Lu, and K. Claxton. Expected Value of Sample Information Calculations in Medical Decision Modeling. *Medical Decision Making*, 24:207–227, 2004.
[8] N. Welton, J. Madan, D. Caldwell, T. Peters, and A. Ades. Expected value of sample information for multi-arm cluster randomized trials with binary outcomes. *Medical Decision Making*, 34(3):352–365, 2014.

[9] A. Brennan and S. Kharroubi. Expected value of sample information for Weibull survival data. *Health Economics*, 16(11):1205–1225, 2007.

[10] M. Strong, J. Oakley, A. Brennan, and P. Breeze. Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample A Fast Nonparametric Regression-Based Method. *Medical Decision Making*, 35(5):570–583, 2015.

[11] N. Menzies. An efficient estimator for the expected value of sample information. *Medical Decision Making*, 36(3):308–320, 2016.

[12] H. Jalal, J. Goldhaber-Fiebert, and K. Kuntz. Computing expected value of partial sample information from probabilistic sensitivity analysis using linear regression metamodeling. *Medical Decision Making*, 35(5):584–595, 2015.

[13] H. Jalal and F. Alarid-Escudero. A Gaussian Approximation Approach for Value of Information Analysis. *Medical Decision Making*, 38(2):174–188, 2018.

[14] A. Brennan and S. Kharroubi. Efficient computation of partial expected value of sample information using Bayesian approximation. *Journal of Health Economics*, 26(1):122–148, 2007.

[15] A. Heath, I. Manolopoulou, and G. Baio. Efficient Monte Carlo Estimation of the Expected Value of Sample Information using Moment Matching. *Medical Decision Making*, 38(2):163–173, 2018.

[16] A. Heath and G. Baio. Calculating the Expected Value of Sample Information Using Efficient Nested Monte Carlo: A Tutorial. *Value in Health*, 21(11):1299–1304, 2018.

[17] A. Heath, I. Manolopoulou, and G. Baio. Bayesian Curve Fitting to Estimate the Expected Value of Sample Information using Moment Matching Across Different Sample Sizes. Accepted to *Medical Decision Making*, in press(·):–, 2018.

[18] D. Meltzer, T. Hoomans, J. Chung, and A. Basu. Minimal modeling approaches to value of information analysis for health research. *Medical Decision Making*, 31(6):E1–E22, 2011.

[19] A. Stinnett and J. Mullahy. Net health benefits a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making*, 18(2):S68–S80, 1998.

[20] J. Felli and G. Hazen. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making*, 18:95–109, 1998.

[21] D. Coyle and J. Oakley. Estimating the expected value of partial perfect information: a review of methods. *The European Journal of Health Economics*, 9(3):251–259, 2008.

[22] A. Heath, I. Manolopoulou, and G. Baio. A Review of Methods for Analysis of the Expected Value of Information. *Medical Decision Making*, 37(7):747–758, 2017.

[23] G. Baio and P. Dawid. Probabilistic sensitivity analysis in health economics. *Statistical Methods in Medical Research*, 24(6):615–634, 2011.

[24] EUnetHTA. Methods for health economic evaluations: A guideline based on current practices in Europe - second draft, 29th September 2014.

[25] Department of Health and Ageing. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee: Version 4.3, 2008.

[26] Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada [3rd Edition]., 2006.

[27] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[28] Donald B Rubin. Using the SIR algorithm to simulate posterior distributions. *Bayesian Statistics*, 3:395–402, 1988.
[29] W. Sullivan, M. Hirst, S. Beard, D. Gladwell, F. Fagnani, J. Bastida, Cl Phillips, and W. Dunlop. Economic evaluation in chronic pain: a systematic review and de novo flexible economic model. *The European Journal of Health Economics*, 17(6):755–770, 2016.

[30] J. Goldhaber-Fiebert and H. Jalal. Some health states are better than others: using health state rank order to improve probabilistic analyses. *Medical Decision Making*, 36(8):927–940, 2016.

[31] R. Ikenberg, N. Hertel, Andrew M., M. Obradovic, G. Baxter, P. Conway, and H. Liedgens. Cost-effectiveness of tapentadol prolonged release compared with oxycodone controlled release in the UK in patients with severe non-malignant chronic pain who failed 1st line treatment with morphine. *Journal of Medical Economics*, 15(4):724–736, 2012.

[32] S. Gates, M. Williams, E. Withers, E. Williamson, S. Mt-Isa, and S. Lamb. Does a monetary incentive improve the response to a postal questionnaire in a randomised controlled trial? The MINT incentive study. *Trials*, 10(1):44, 2009.

[33] F. Alarid-Escudero, R. MacLehose, Y. Peralta, K. Kuntz, and E. Enns. Nonidentifiability in model calibration and implications for medical decision making. *Medical Decision Making*, 38(7):810–821, 2018.

[34] Human Mortality Database. Available at [www.mortality.org](http://www.mortality.org) or [www.humanmortality.de](http://www.humanmortality.de) (data downloaded on [2019-01-22]), 2019.

[35] R. Hogg and A. Craig. Introduction to mathematical statistics. (5th edition). Englewood Hills, New Jersey, 1995.

[36] M. Strong, J. Oakley, and A. Brennan. Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample A Nonparametric Regression Approach. *Medical Decision Making*, 34(3):311–326, 2014.

## A Inputs for the Chemotherapy Model

| Model Input                                                                 | Distribution     | 1st Prior Parameter | 2nd Prior Parameter | Previous Data                                                                 |
|----------------------------------------------------------------------------|------------------|---------------------|---------------------|-----------------------------------------------------------------------------|
| πₐ - Probability of adverse events                                        | Beta             | 1                   | 1                   | Number of adverse events                                                    |
| ρ - Reduction in adverse events with treatment                            | Normal           | Mean: 0.65          | Precision: 100      | No                                                                          |
| q - QoL weight with no adverse events                                     | Beta             | 18.23               | 0.372               | Number of hospitalizations                                                   |
| Γ₁ - Probability of hospitalisation                                       | Beta             | 1                   | 1                   | Number of deaths                                                            |
| Γ₂ - Probability of death                                                 | Beta             | 1                   | 1                   |                                                                             |
| γ₁ₐ - Daily transition probability to hospital                            | Beta             | 5.12                | 6.26                | No                                                                          |
| γ₂ₐ - Daily probability of death                                          | Beta             | 3.63                | 6.74                |                                                                             |
| λ₀ - Daily probability of recovery from home care                         | Beta             | 8.33                | 0.13                |                                                                             |
| Cost of death                                                             | LogNormal        | 8.77                | 0.15                |                                                                             |
| Cost of hospital care                                                     | LogNormal        | 8.77                | 0.15                |                                                                             |
| QoL weight for hospital care                                              | Beta             | 5.75                | 5.75                |                                                                             |
| QoL weight for hospital care                                              | Beta             | 0.87                | 3.47                |                                                                             |

Table 3: The prior specification for the parameters underlying the Chemotherapy example including the distributional assumption and its parameters. Unless specified, the parameters are specified in the order used in the JAGS language for Bayesian updating. We indicate whether the stated prior is combined with data in the probabilistic health economic model.