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

Anna Heath[1,2,3], Natalia R. Kunst[4,5,6,7], Christopher Jackson[8], Mark Strong[9], Fernando Alarid-Escudero[10], Jeremy D. Goldhaber-Fiebert[11], Gianluca Baio[3], Nicolas A. Menzies[12], Hawre Jalal[13], and on behalf of the Collaborative Network for Value of Information (ConVOI)

[1]The Hospital for Sick Children
[2]University of Toronto
[3]University College London
[4]University of Oslo
[5]Yale University School of Medicine
[6]Amsterdam UMC
[7]LINK Medical Research
[8]MRC Biostatistics Unit, University of Cambridge
[9]School of Health and Related Research, University of Sheffield
[10]Center for Research and Teaching in Economics (CIDE)
[11]Stanford Health Policy, Centers for Health Policy and Primary Care and Outcomes Research, Stanford University
[12]Harvard TH Chan School of Public Health
[13]University of Pittsburgh

Abstract

Background: 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. Recently, several more efficient EVSI approximation methods have been developed. However, these approximation methods have not been compared and therefore their comparative performance across different examples has not been explored.

Methods: We 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.

Results: 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.

Conclusions: As all the evaluated methods gave estimates similar to those given by traditional Monte Carlo, we suggest that EVSI can now be efficiently computed with confidence in realistic examples. No systematically superior EVSI computation method exists as the properties of the different methods depend on the underlying health economic model, data generation process and user expertise.

Introduction

The Expected Value of Sample Information (EVSI) [1, 2] 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 [3], 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 [4, 5]. Among these is the high computational cost associated with obtaining precise estimates of EVSI in real-world scenarios using nested Monte Carlo (MC) sampling [6]. This computational burden is increased further when we aim to identify the optimal research study (i.e., the study with the highest ENBS) and must compute EVSI for multiple trial designs [7, 8]. High performance computing can be used to overcome some of these barriers, but this requires additional programming skills and increases the complexity of the analysis.

Several methods have been developed to overcome these computational barriers and, thus, use EVSI for research prioritization and trial design optimization [9–20]. However, as many of these methods have been developed concurrently, they have not been compared. Additionally, these EVSI estimation methods have mostly been evaluated using health economic models and trial designs chosen for computational convenience rather than to reflect real-world decision making. Thus, the accuracy of the majority of these methods has not been assessed in, for example, multi-state models combined with survival or quality-of-life outcomes, a setting that would occur frequently in real-life health economic modelling.

We aim, therefore, to address these gaps by comparing the relative performance of EVSI calculation methods across three realistic health economic models and trial designs to gain a greater understanding of their behaviour in practice. We will evaluate the accuracy of the EVSI estimation methods across our three examples and the computational time required to obtain these estimates. These examples exhibit features that reflect real-world trial design and may influence the behaviour of these EVSI estimation methods 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.

Our comparison is considers four recent calculation methods that impose very limited restrictions on the structure of the underlying health economic model, the number of evidence sources and the study design. These methods are the Regression Based (RB) method developed by Strong et al. [12], the Importance Sampling (IS) method developed by Menzies [13], the Gaussian Approximation (GA) method developed by Jalal and Alarid-Escudero [15] (extending a method proposed in Jalal et al. [14]), and the Moment Matching (MM) method developed by Heath et al. [17–19]. These methods are based on different approaches and assumptions, but they all provide EVSI estimates with a smaller computational burden compared to the nested MC sampling methods whilst retaining accuracy.

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 some measure of benefit (which is usually net monetary benefit or net health benefit [21]). We characterize a health economic model as a function that takes the vector of parameters $\boldsymbol{\theta}$ as an input, 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 $p(\boldsymbol{\theta})$. 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 $\boldsymbol{\theta}$. Uncertainty about $\boldsymbol{\theta}$ induces uncertainty about the net benefit for each treatment $t = 1, \ldots, T$. We denote the net benefit for treatment t given parameters $\boldsymbol{\theta}$ as $\mathrm{NB}_t^{\boldsymbol{\theta}}$. 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 we are interested in collecting additional information about a subset of the model parameters $\boldsymbol{\phi}$. In this setting, $\boldsymbol{\theta}$ is split into two sets of parameters $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\psi})$, where $\boldsymbol{\psi}$ are all the remaining model parameters not in $\boldsymbol{\phi}$. 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 $\boldsymbol{\phi}$ (assuming risk neutrality) is equal to the Expected Value of Partial Perfect Information (EVPPI) [22–24]. This is given by

$$\mathrm{EVPPI} = \mathrm{E}_{\boldsymbol{\phi}} \left[\max_t \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{\phi}} \left[\mathrm{NB}_t^{\boldsymbol{\theta}} \right] \right] - \max_t \mathrm{E}_{\boldsymbol{\theta}} \left[\mathrm{NB}_t^{\boldsymbol{\theta}} \right]. \tag{1}$$

In this paper, EVSI is defined as the value of collecting additional data, denoted $\boldsymbol{X}$, to inform $\boldsymbol{\phi}$,

2

where ϕ could be equal to θ if the additional data updates all the underlying model parameters. By assuming that X directly updates ϕ only, we implicitly state that $X \perp\!\!\!\perp \psi \mid \phi$. This implies that EVSI is bounded above by the EVPPI for ϕ. If these data X had been collected and observed to have a value x, they would be combined with the current evidence to generate an updated distribution for ϕ, $p(\phi \mid x)$. 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 ϕ. 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

$$\text{EVSI} = \text{E}_{\boldsymbol{X}} \left[\max_t \text{E}_{\boldsymbol{\theta}|\boldsymbol{X}} \left[\text{NB}_t^{\boldsymbol{\theta}} \right] \right] - \max_t \text{E}_{\boldsymbol{\theta}} \left[\text{NB}_t^{\boldsymbol{\theta}} \right], \tag{2}$$

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

Calculation Methods for EVSI

It is rarely possible to compute EVSI analytically as the net benefit is often a complex function of $\boldsymbol{\theta}$. Additionally, it is challenging to compute the expectation of a maximum analytically as required in the first term of equation (5). Therefore, a range of methods have been developed to approximate EVSI.

Nested Monte Carlo Computations for EVSI

The simplest approximation method [6] computes all the expectations in equation (5) using MC simulation. The second term can be computed by simulating $s = 1, \ldots, S$ parameter values, $\boldsymbol{\theta}_s$, from $p(\boldsymbol{\theta})$. The simulated values are used as inputs to a health economic model to obtain S simulations of the net benefit for each intervention, denoted $\text{NB}_t^{\boldsymbol{\theta}_s}$. Note that this process is required to perform a "probabilistic analysis" (PA) [25], used to assess the impact of parametric uncertainty on the decision uncertainty, which is mandatory in various jurisdictions [26–28]. The average of $\text{NB}_t^{\boldsymbol{\theta}_1}, \ldots, \text{NB}_t^{\boldsymbol{\theta}_S}$ for each intervention can be computed and $\max_t \text{E}_{\boldsymbol{\theta}} \left[\text{NB}_t^{\boldsymbol{\theta}} \right]$ is estimated by the maximum of these means.

The first term in equation (5) is more complex to compute by simulation. Firstly, S datasets $\boldsymbol{X}_s$ must be generated conditional on the simulated $\boldsymbol{\theta}_s$ from the assumed sampling distribution $p(\boldsymbol{X} \mid \boldsymbol{\theta}_s)$. For each $\boldsymbol{X}_s$, we simulate R values from the updated distribution of the model parameters $p(\boldsymbol{\theta} \mid \boldsymbol{X}_s)$. 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 $\text{E}_{\boldsymbol{\theta}|\boldsymbol{X}} \left[\text{NB}_t^{\boldsymbol{\theta}} \right]$ for $t = 1, \ldots, T$. The maximum of these simulated means is then selected for each $\boldsymbol{X}_s$. 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 net benefit associated with each intervention t using a smaller simulation burden. We denote the expectation of the net benefit, conditional on data $\boldsymbol{X}$, as

$$\mu_t^{\boldsymbol{X}} = \text{E}_{\boldsymbol{\theta}|\boldsymbol{X}} \left[\text{NB}_t^{\boldsymbol{\theta}} \right].$$

In a similar manner, we also denote the expectation of the net benefit, conditional on some value of the parameters of interest ϕ, as

$$\mu_t^{\phi} = \text{E}_{\boldsymbol{\theta}|\phi} \left[\text{NB}_t^{\boldsymbol{\theta}} \right].$$

Finally, we can increase the numerical stability of the following approximation methods by working in terms of the incremental net benefit or loss, defined, without loss of generality, as $\text{INB}_t^{\boldsymbol{\theta}} = \text{NB}_t^{\boldsymbol{\theta}} - \text{NB}_1^{\boldsymbol{\theta}}$ for $t = 2, \ldots, T$. This is because we must estimate $\mu^{\boldsymbol{X}}$ for each of the $T - 1$ incremental net benefit options rather than the T net benefits. EVSI can then be estimated by setting $\mu_1^{\boldsymbol{X}} = 0$ and calculating

$$\widehat{\text{EVSI}} = \frac{1}{S} \sum_{s=1}^{S} \max_t \mu_t^{\boldsymbol{X}_s} - \max_t \frac{1}{S} \sum_{s=1}^{S} \mu_t^{\boldsymbol{X}_s}, \tag{3}$$

3

where $\mu_t^{\boldsymbol{X}_s}$ is the estimated posterior expectation for the dataset $\boldsymbol{X}_s$, $s = 1, \ldots, S$.

Regression Based Method

The RB method [12] estimates EVSI by fitting $T - 1$ regression models. For each $t = 2, \ldots, T$, the simulated values of the incremental net benefit are the 'response' variable and a low-dimensional summary of the simulated dataset $\boldsymbol{X}$ is the 'predictor' variable(s) [12]. This low-dimensional summary for $\boldsymbol{X}$ should reflect how the data would be summarized if the study were to go ahead and must be computed for each simulated dataset $\boldsymbol{X}_s$. The fitted values from this regression model are then used to estimate $\mu_t^{\boldsymbol{X}}$. EVSI is then estimated directly from these estimates of $\mu_t^{\boldsymbol{X}}$ using equation (6).

Importance Sampling Method

Initially, two IS methods were proposed to estimate EVSI [13], the most accurate of these estimates $\mu_t^{\boldsymbol{X}}$ by reweighting simulations of $\mu_t^{\boldsymbol{\phi}}$. This reweighting is based on the *likelihood* of observing a simulated dataset $\boldsymbol{X}_s$ conditional on different values for $\boldsymbol{\phi}$. The term likelihood is used in the statistical sense and is equal to $p(\boldsymbol{X} \mid \boldsymbol{\phi})$.

This method simulates S future datasets $\boldsymbol{X}_s$ from $p(\boldsymbol{X} \mid \boldsymbol{\phi}_s)$. The likelihood for *every* simulated vector for $\boldsymbol{\phi}$ is then calculated conditional on $\boldsymbol{X}_s$. For the sample $\boldsymbol{X}_s$, $\mu_t^{\boldsymbol{X}_s}$ is estimated as the weighted average of $\mu_t^{\boldsymbol{\phi}}$, weighted by the likelihood of the dataset $\boldsymbol{X}_s$. Thus, the IS method is an example of importance sampling [29, 30] and performs $(T - 1)S^2$ likelihood calculations for each EVSI estimate. EVSI is estimated from equation (6) based on the estimate of $\mu_t^{\boldsymbol{X}_s}$ for each future sample.

Gaussian Approximation Methods

The GA method [14, 15] fits a "linear" metamodel[1], a secondary model that captures the relationship between the simulated incremental net benefit values, as the response variable, and simulations for $\boldsymbol{\phi}$, as the predictor variables. Each term of the linear metamodel is then rescaled based on a Gaussian-Gaussian Bayesian updating approach to estimate its 'posterior' expectation across different future datasets $\boldsymbol{X}$. These estimated distributions are then recombined using the coefficients of the linear metamodel to estimate $\mu_t^{\boldsymbol{X}}$ and compute EVSI.

For a proposed future data collection strategy of size N, the rescaling factor for each term of the linear metamodel is equal to

$$\frac{N}{N + N_0},$$

where N_0 is known as the prior effective sample size [31]. Essentially, N_0 represents the number of independent observations who would be required to generate the amount of evidence in the prior. In some prior-likelihood pairs, N_0 can be obtained analytically. In other settings, Jalal *et al.* suggest two estimation methods for N_0. Firstly, if the data $\boldsymbol{X}$ can be summarized using a summary statistic $W(\boldsymbol{X})$, then N_0 can be computed as a function of the variance of $W(\boldsymbol{X})$. Secondly, if a suitable statistic cannot be derived, then nested posterior sampling can be used to estimate N_0. In this method, S future datasets $\boldsymbol{X}_s$, $s = 1, \ldots, S$ are simulated. Each of these samples is used to update the information about the model parameters $p(\boldsymbol{\theta} \mid \boldsymbol{X}_s)$, typically using R simulations and computing the mean for $\boldsymbol{\phi}$. The *variance* of the mean for $\boldsymbol{\phi}$, across different samples $\boldsymbol{X}_s$, is then used to estimate N_0. Computationally, this nested sampling method to compute N_0 is relatively 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.

Moment Matching Method

The MM method [18, 19] combines the simulations $\mu_t^{\boldsymbol{\phi}}$ 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 [19]. Thus, EVSI is estimated with $Q \times R$ health economic model runs.

The MM method uses nested MC sampling to estimate the *variance* of the incremental net benefit for different future datasets. These estimated variances rescale simulations of $\mu_t^{\boldsymbol{\phi}}$ for $t = 2, \ldots, T$ to approximate simulations of $\mu_t^{\boldsymbol{X}}$ which can be used to estimate EVSI using equation (6). The MM

---

[1]A linear metamodel is required for this method. However, non-linear functions of $\boldsymbol{\phi}$ can be defined and combined linearly to account for flexible relationships between the incremental net benefit and the parameters $\boldsymbol{\phi}$.

method only requires a single nested simulation procedure to estimate EVSI across different sample sizes of the future trial [19].

Of note, this comparison has been restricted to these four methods as the alternative EVSI methods typically place restrictions on the structure of the underlying health economic model, the number of evidence sources and/or the study design [9–11, 11, 20]. These restrictions limit the applicability of these methods. Some of these methods make assumptions about the distribution of the parameters and the study data to ensure that the prior and posterior model parameter distributions take the same form (conjugacy). This allows for computationally efficient EVSI estimation. The minimal modelling approach [20] assumes that a comprehensive clinical trial is available to inform EVSI estimation and, thus, restricts the required data sources. These methods are competitive in terms of computational time and accuracy compared to the methods presented in this review but are restricted to the settings in which they are relevant which limits their general purpose application.

Case Studies

These EVSI methods are applied to three case studies designed to explore different trial designs and health economic models. These designs and models were chosen to assess the accuracy of these EVSI estimation methods in settings that are reflective of real-world decision making. The first case study is a stylized Chemotherapy side effects 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 Model for Chemotherapy Side Effects

This model was presented in Heath and Baio [18] 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 π_0 and ρ 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, γ_1 and γ_2 denote the constant probability of requiring hospital care and dying, respectively, and λ_1 and λ_2 denote the constant probability of recovery given that an individual is not or is admitted into hospital, respectively. The cycle length is 1 day, the time horizon is 15 days, and we assume that only hospitalised patients will die. 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. PA distributions for the model parameters are informed using previous data or defined using expert opinion with all distributional assumptions given in the supplementary material.
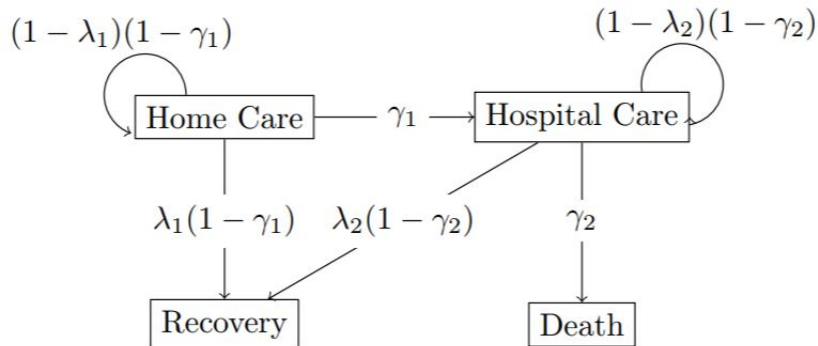

Figure 1: A four state Markov model used to model the health economic impact of adverse events from a chemotherapy treatment.

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 with 150 patients per arm.

To define the sampling distribution for the six outcomes, we model the number of patients who experience adverse events using binomial distributions conditional on π_0 and ρ;

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

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

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

Finally, recovery time for patients who experience adverse events but recover whilst remaining at home is modelled with an exponential distribution conditional on the transition probability λ_1,

$$T_{HC}^i \sim Exponential(\eta_1)$$

for each individual who remains at home whilst recovering from adverse events $i = 1, \ldots, X_{AE_0} + X_{AE_1} - X_{Hosp}$ and $\eta_1 = -\log(\lambda_1)$. The recovery time for every patient ($j = 1, \ldots, X_{Hosp} - X_{Death}$) who recovers in hospital is modelled as and exponential distribution conditional on λ_2

$$T_H^j \sim Exponential(\eta_2)$$

with $\eta_2 = -\log(\lambda_2)$. Exponential distributions were used to model the recovery time as we assumed a constant transition probability in the Markov model.

Case Study 2: A Model for Chronic Pain

This example uses a cost-effectiveness model developed by Sullivan et al. [32], and extended in Heath *et al.* [19], 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. Patients initially receive treatment for chronic pain and can either experience adverse events or not. Patients may then withdraw from this initial treatment due to adverse events or lack of efficacy. Following this, they can be offered an alternative therapy or withdraw completely from treatment. Patients may experience adverse events from this second-line treatment and can either withdraw from this treatment due to these adverse events or lack of efficacy. If patients withdraw from this second-line treatment, they can receive further treatment or discontinue, both considered absorbing states as the model does not include a death state.

Initially, patients can either be offered morphine or an innovative treatment, and our model evaluates the cost-effectiveness of the innovative treatment. If patients require a second-line treatment, they are offered oxycodone. Thus, the only difference between the two treatment 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 [32] where the distributions for the PA are gamma for costs and beta for probabilities and utilities. In the original publication of this model, the means of these distributions were informed by relevant studies identified following a literature review and the standard error of these mean estimates was taken as 10% of the underlying mean estimate. This paper computed the per-person lifetime EVSI, 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 for patients who withdraw from the first-line treatment due to lack of efficacy. The individual level variability in these two QoL weights is modelled, for simplicity, as independent beta sampling distributions although the assumption of independence may be invalid [33]. The population level mean QoL weight, i.e., the mean of the sampling distribution for the QoL weights, is defined as the value of those two health states in the Markov Model. The standard deviations of the individual level sampling 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 [34][2]. We compute EVSI for trials enrolling

---

[2]This sampling distribution for the data causes some minor issues for the Gibbs sampling procedure used in the JAGS program for Bayesian updating.

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 [35]. 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 Screening

This example uses a health economic model developed by Alarid-Escudero *et al.* [36] to evaluate a screening strategy for colorectal cancer (CRC) and pre-cancerous lesions known as adenomas. The 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 modelled using a Weibull hazard conditional on age

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

where λ_1 and g are the shape and scale parameters of the Weibull distribution and a is the age of the patient. Model parameters are chosen to reflect the prevalence of adenomas seen in the literature. To determine the PA distribution for the model parameters g and λ_1, the level of uncertainty in the observed prevalence was characterized and the parameters were re-estimated across the distribution of prevalence.

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. When the model is initiated, some members of the general population have undiagnosed adenomas and early stage CRC.

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 [37], were used to generate study subjects with an age distribution representative of the general population, with study enrolment 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 $p(a_i)$

$$X_i \sim Ber\left(p\left(a_i\right)\right).$$

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

It is challenging to compare the four efficient EVSI estimation methods as their accuracy and computational time are dependent on choices made by the modeller and the computational efficiency of the method implementation. Thus, Table 3 outlines the simulation choices that were made for the case studies in this paper. We chose these assumptions to reflect realistic choices that would be made by modellers calculating EVSI in practice. These choices all aim to estimate EVSI with a reasonable level of precision, while keeping the computation time manageable. Based on these choices, we compared the speed and accuracy of each method, and identified their relative advantages and challenges in practice.

The Chemotherapy side effects and Chronic pain models had low computational cost and, thus, we selected a large PA simulation size S as VoI methods require an accurate representation of the decision

uncertainty [38]. The CRC screening model had greater computational cost and thus, we used a smaller simulation size to reflect choices that would be made in practice. In all case studies, we used all the PA simulations to calculate EVPPI and estimate μ_t^ϕ.

For the MC nested simulation, we used all the PA simulations for the outer simulation loop. To accurately characterise the PA distribution for the inner loop, we use the same size as the outer loop. This gives a very high simulation burden but leads to accurate results for EVSI that can be compared with the novel computation methods.

Simulation Choices	Case Study		
	Chemotherapy side effects (1)	Chronic pain (2)	CRC screening (3)
Initial PA size (S)	100,000	100,000	5,000
Number of μ_t^ϕ simulations from EVPPI calculation	100,000	100,000	5,000
Nested simulation outer loop size (S)	100,000	100,000	NA
Nested simulation inner loop size (R)	100,000	100,000	NA
RB simulation size	100,000	100,000	5,000
IS simulation size	20,000	5,000	2,500
GA N_0 computation method	nested posterior sampling	nested posterior sampling	nested posterior sampling
GA N_0 estimation outer loop size	1,000	1,000	5,000
GA N_0 estimation inner loop size	10,000	10,000	5,000
GA N_0 estimation future sample size	30	40	40
MM outer loop size (Q)	50	50	50
MM inner loop size (R)	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.

For all case studies, the RB method uses all the PA simulations as regression modelling requires larger simulation sizes to accurately estimate EVSI. The computational cost of using all the simulations is minimal. In contrast, we used a smaller simulation size for the IS method as it has a higher computational cost compared to the other methods that increases proportional to S^2 and we wanted to ensure a comparable analysis in terms of computation time.

For the GA method, we used nested posterior sampling to estimate the prior effective sample size for all three case studies. N_0 is estimated using a proposed future sample with a given sample size but only needs to be computed once to estimate the EVSI across sample size. As posterior updating is slower for larger sample sizes, we can reduce the computational cost of estimating N_0 by using a small sample size for the proposed sample $\boldsymbol{X}$. However, as the estimation of N_0 also relies on a Gaussian approximation, the sample size of $\boldsymbol{X}$ should be sufficiently large to assume normality. Thus, we selected future sample sizes for the GA updating of around 30 (or 40 to adjust for missingness) to ensure the assumption of normality holds but to decrease computational complexity. The simulation sizes for the nested simulation were chosen to balance accuracy and computation time. Specifically, we chose more simulations for the CRC screening model as the posterior of the parameters for a Weibull distribution is highly correlated and therefore challenging to estimate using Gibbs sampling [39].

Finally, Heath and Baio determined that $Q = 50$ is sufficient to estimate EVSI using MM method, provided the inner loop size R is sufficient to capture the posterior for the model parameters [19]. Thus, we selected relatively large inner loop sizes to capture the posterior distribution of the model parameters, with the CRC screening model simulation size reduced due to the computational complexity of that example.

In general, larger PA simulation sizes lead to increased computational time and accuracy for all methods. The improvements associated with larger sample sizes change depending on the method and the case study. For example, the RB and GA methods require a greater number of PA simulations if

there are a large number of outcomes in the proposed trial as the regression model or the metamodel increases in complexity. Elsewhere, the MM method requires more PA simulations if the underlying EVSI is small compared to the overall value of resolving parametric uncertainty. Nonetheless, we believe that the choices highlighted in Table 3 represent a fair comparison of these methods that supports the use of these methods based on modelling choices that could be implemented in practice.

To characterise uncertainty in the EVSI estimation procedure for each method, we computed a standard error for the EVSI estimates by recomputing the EVSI 200 times, based on the same PA simulations for the first two case studies. The computational time for the four recent approximation methods is presented based on computations undertaken on a computer with an i7 Intel processor with 16 GB of RAM in R version 3.5.1. The computation time for the nested MC computation is based on the total computation time required across 32 cores on the Hospital for Sick Children's High Performance Computing environment. Code to undertake the computations in this paper is available from GitHub at `https://github.com/convoigroup/EVSI-in-practice`.

Results

Case Study 1: Chemotherapy Side Effects

Figure 6 displays the 95% central intervals for the four fast 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. The 95% central interval for the RB and MM methods contain the 'true' value, represented by the nested MC EVSI. In this example, however, the MM estimate is associated with substantial variability compared to the other methods as the EVSI is smaller compared to the overall EVPI.
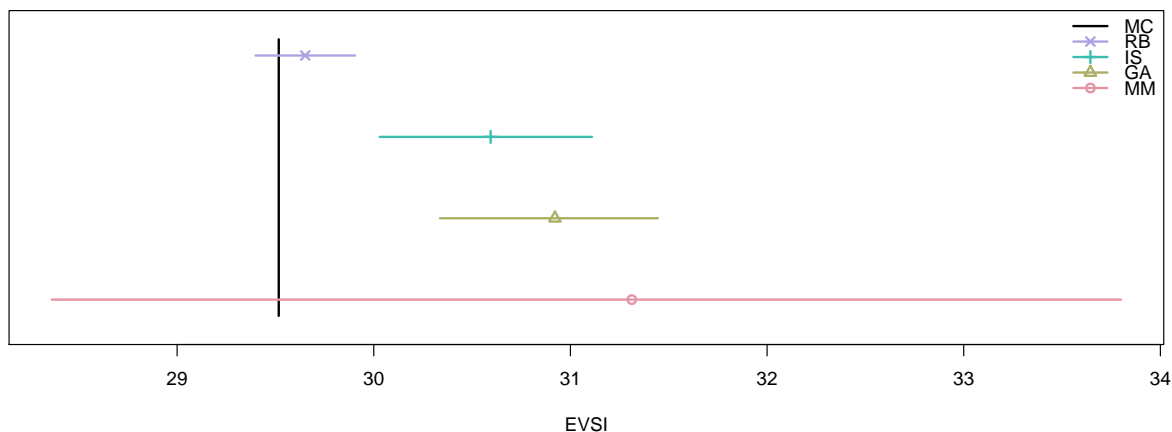

Figure 2: The mean per-person EVSI estimates, across 200 simulated estimation procedures, for the five methods under consideration for the Chemotherapy side effects example with a future sample size of 150 and willingness-to-pay of £30,000. The considered methods are: the nested Monte Carlo estimator (MC), the Regression Based method (RB), the Importance Sampling method (IS), the Gaussian approximation method (GA) and the Moment Matching method (MM). 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 RB and GA 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 7 shows that the 95% central intervals for the MM and the IS methods contain the nested MC estimate for all sample sizes. In addition to this, EVSI calculated by all four methods increase as the sample size increases, representing that the more information is collected, the greater its value. EVSI estimates should also remain below EVPPI (marked as a dashed line on Figure 7), which represents an

upper limit on EVSI. The EVSI estimates are also relatively close to the nested MC estimate. The RB method produced the shortest 95% central intervals while the three alternatives are relatively comparable. Note that the IS estimate is based on a smaller PA simulation size but still offers similar variability compared to the other methods.
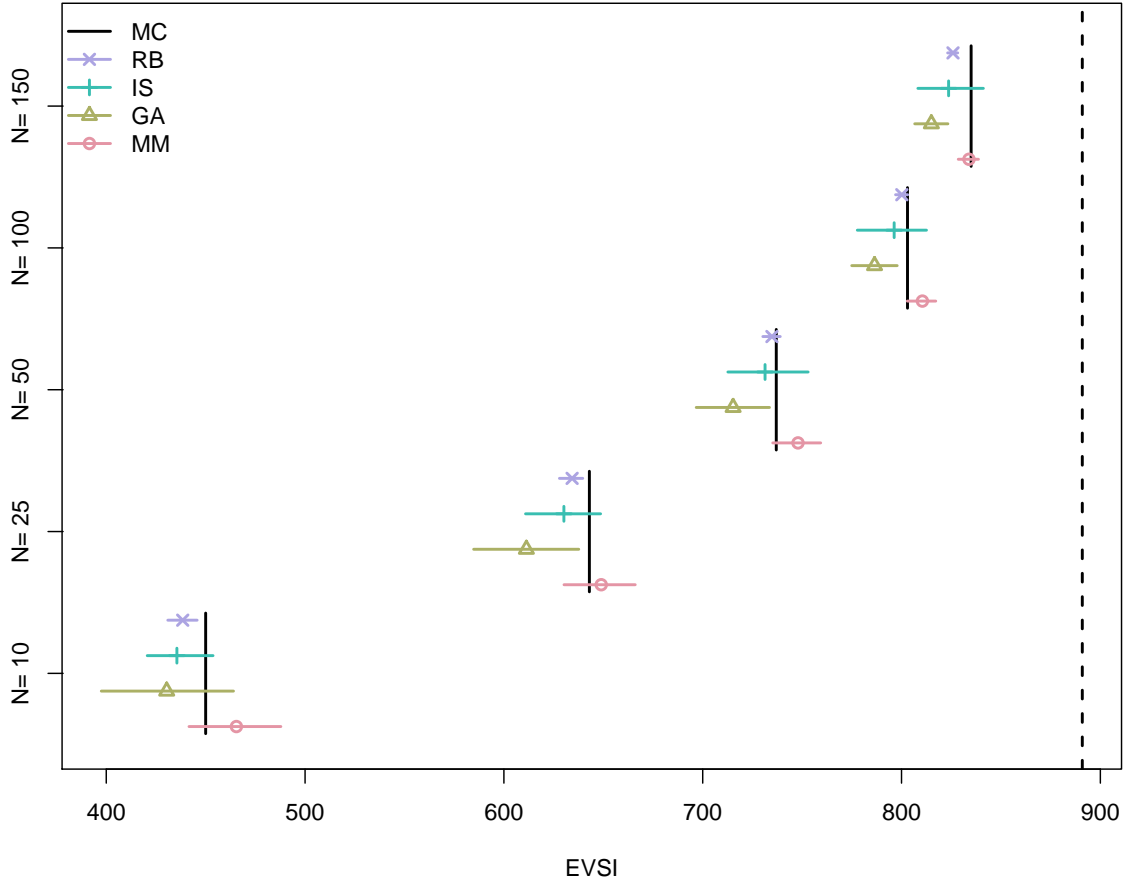

Figure 3: The mean EVSI estimates, across 200 simulated estimation procedures, for the five methods under consideration for the Chronic pain example. EVSI was calculated across 5 different sample sizes for the future trial. The considered methods are: the nested Monte Carlo estimator (MC), the Regression Based method (RB), the Importance Sampling method (IS), the Gaussian approximation method (GA) and the Moment Matching method (MM). 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 RB method is the geometric mean of $\boldsymbol{X}$ and $1 - \boldsymbol{X}$. These statistics are sufficient to estimate the model parameters of the beta distribution and were derived using the Fisher-Neymann factorization theorem [40]. Summarizing $\boldsymbol{X}$ using the arithmetic mean and variance gives incorrect EVSI estimates for this case study as these statistics are not sufficient. Low-dimensional sufficient statistics should be used, when available, to estimate EVSI if the proposed analysis following the trial will use Bayesian methods to incorporate the additional information into the evidence base of the health economic model.

Case Study 3: Colorectal Cancer Screening

Figure 8 demonstrates EVSI estimates (y-axis) for the CRC screening model across the considered sample sizes (x-axis, on the log-scale but marked on the natural scale). We can see that the four EVSI calculation methods give a broad consensus. 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 note that the four methods give similar results, we cannot assert that these EVSI estimates are 'correct.'

Of note, for a sample size of 1,500, the IS EVSI estimate is incorrect. This is because the likelihood tends to 0 for large sample sizes making the weights in the weighted sum challenging to approximate. Furthermore, the IS method slightly over-estimates the EVSI for sample sizes between 500 and 1,000. This is because we only use a subset of the PA 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 PA simulations.
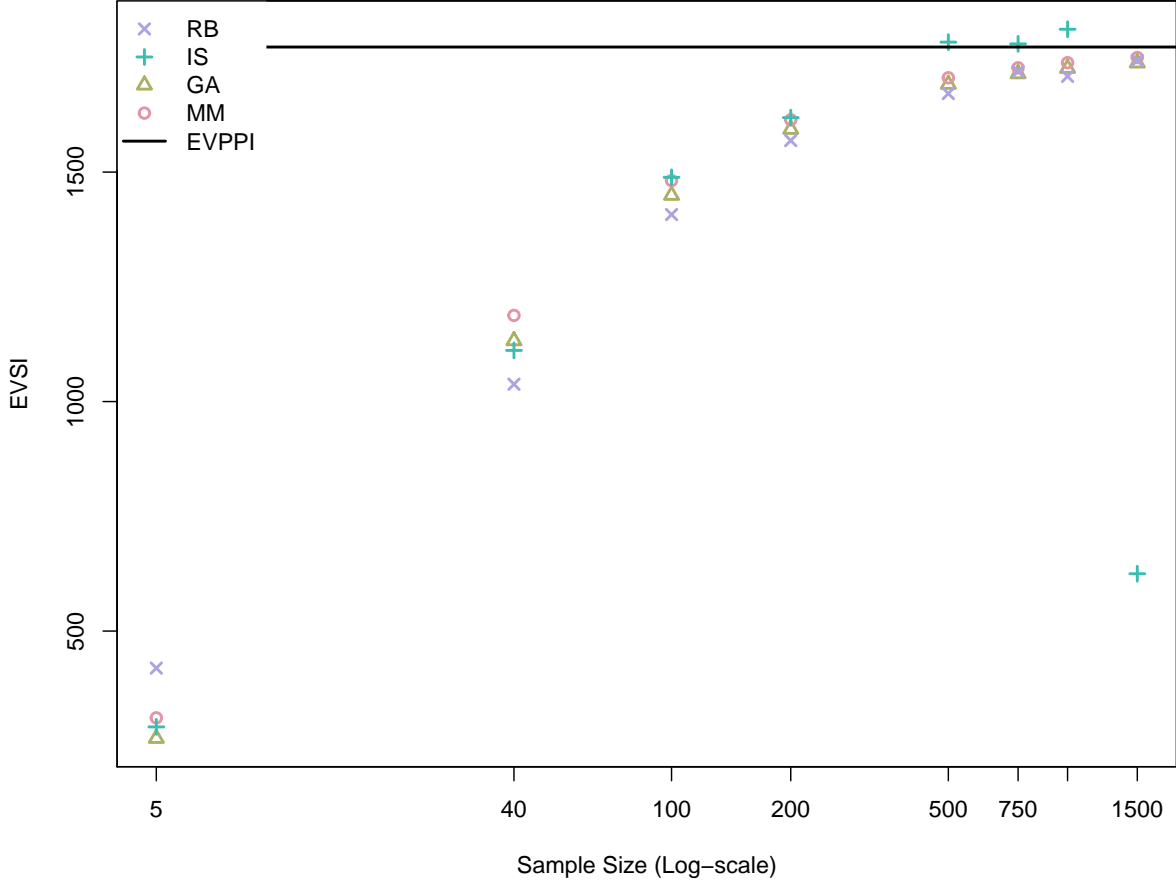
Figure 4: EVSI estimates for the four methods under consideration for the CRC screening model. The considered methods are: the Regression Based method (RB), the Importance Sampling method (IS), the Gaussian approximation method (GA) and the Moment Matching method (MM). EVSI is calculated for 9 different 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 [41], is included as a black line on this Figure.

Computational Time

Table 4 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 screening model meant that it was not computationally feasible to use the nested Monte Carlo method.

For the first two case studies, the MM method has the lowest computation time as the underlying health economic model is fast. The MM method also estimates EVSI across multiple sample sizes simultaneously which improves the computational time for the Chronic pain example compared to the RB, and IS methods. For these two examples, the computation time required to fit an accurate regression model is relatively high, increasing the computation time for the RB method. The GA 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 GA method would offer computational savings on the RB and IS methods. For the

Chemotherapy side effects example, the IS 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 PA simulations are used, the computation time is greater than 2 hours. For the Chronic pain example, the IS method is noticeably slower as the computation time for the likelihood increases when the proposed sample size of $\boldsymbol{X}$ is larger.

Case Study	Computational Time (mins)				
	Nested MC	RB	IS	GA	MM
1: Chemotherapy side effects	37646	5.35	4.56	8.20	2.01
2: Chronic pain	223200	12.05	86	22.27	2.46
3: CRC screening	*	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 GA 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 RB method, $\boldsymbol{X}$ is summarized by finding the maximum likelihood estimates (MLE) for g and λ_1 that must be estimated, using relatively slow computational optimization procedures, for each sample $\boldsymbol{X}_s$, $s = 1, \ldots, S$ and sample size. Thus, estimating the summary statistics is slow in this case study. The MM method is more computationally expensive as the underlying probabilistic sensitivity analysis for the CRC screening health economic model is expensive and must be rerun $Q \times S = 250,000$ times to compute EVSI. The computational time of the IS method is similar to the previous case studies.

Discussion

The paper uses three case studies to compare the accuracy and computational time for four novel methods for approximating EVSI as no previous head-to-head comparison had been done. Additionally, we assessed the performance of these methods using health economic models that were designed to cover a number of different trial designs, interventions and health economic model structures. Thus, we assessed the comparative performance of these efficient EVSI estimation methods across a number of scenarios that may make the EVSI estimation more challenging.

In general, the EVSI estimates are accurate when the underlying assumptions for the respective methods were met and can, thus, be used with confidence in practice. 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. Although, the IS method is generally more computationally intensive than the other methods.

In this analysis, we have not identified an efficient EVSI estimation method that is systematically superior to the alternatives. Specifically, the "optimal" estimation method that trades off accuracy, precision, computational time and ease of implementation will change depending on the health economic model structure, proposed trial design and analyst expertise [42].

Nonetheless, the analysis in this paper has emphasized some distinctions between the methods. Firstly, the RB method is accurate and efficient, provided the analyst can correctly summarize the trial data and fit a regression model, which can be challenging either statistically or computationally. The IS method is accurate but computationally expensive for large PA simulation sizes. The GA method is efficient when estimating EVSI across sample sizes but often requires nested posterior sampling when considering realistic data collection exercises and also relies on an accurate "linear" metamodel. Finally, the MM method is accurate and efficient when the health economic model has a low computation time but becomes more unfeasible as the model run time increases. The GA and MM methods require expertise in Bayesian methods/thinking. The IS method requires the repeated evaluation of the likelihood function, and hence relies on this function being both known and computationally tractable. Beyond these remarks, future research is required to fully articulate the relative strengths and limitations of these methods for different health economic models and trial designs and to support the widespread implementation of these methods.

Contributions

Acknowledgements

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] C. McKenna and K. Claxton. Addressing adoption and research design decisions simultaneously: the role of value of sample information analysis. *Medical Decision Making*, 31(6):853–865, 2011.

[4] 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.

[5] N. Welton and H. Thom. Value of Information We'e Got Speed, What More Do We Need? *Medical Decision Making*, 35(5):564–566, 2015.

[6] A. Brennan, S. Kharroubi, 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.

[7] 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.

[8] 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.

[9] A. Ades, G. Lu, and K. Claxton. Expected Value of Sample Information Calculations in Medical Decision Modeling. *Medical Decision Making*, 24:207–227, 2004.

[10] 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.

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

[12] 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.

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

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

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

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

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

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

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

[30] Donald B Rubin. Using the SIR algorithm to simulate posterior distributions. *Bayesian Statistics*, 3:395–402, 1988.

[31] S. Morita, P. Thall, and P. Müller. Determining the effective sample size of a parametric prior. *Biometrics*, 64(2):595–602, 2008.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] Human Mortality Database. Available at "`www.mortality.org`" or "`www.humanmortality.de`" (data downloaded on [2019-01-22]), 2019.

[38] A. Hatswell, A. Bullement, A. Briggs, M. Paulden, and M. Stevenson. Probabilistic sensitivity analysis in cost-effectiveness models: determining model convergence in cohort models. *PharmacoEconomics*, 36(12):1421–1426, 2018.

[39] W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995.

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

[41] 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.

[42] N.R. Kunst, E. Wilson, F. Alarid-Escudero, G. Baio, A. Brennan, M. Fairley, D. Glynn, J. Goldhaber-Fiebert, C. Jackson, H. Jalal, N. Menzies, M. Strong, H. Thom, and A. Heath. Computing the Expected Value of Sample Information Efficiently: Expertise and Skills Required for Four Model-Based Methods. *arXiv preprint arXiv:1910.03368*, 2019.

Abstract

Background: 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~~ Recently, several 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~~ comparative performance across different examples has not be explored.

Methods: We 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.

Results: 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.

ConclusionConclusions: As all the evaluated methods gave estimates similar to those given by traditional Monte Carlo, we suggest that EVSI can now be efficiently computed with confidence in realistic examples. No systematically superior EVSI computation method exists as the properties of the different methods depend on the underlying health economic model, data generation process and user expertise.

Introduction

The Expected Value of Sample Information (EVSI) [1, 2] 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 [3], 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 [4, 5]. Among these is the ~~computational cost that has typically been~~ high computational cost associated with obtaining precise estimates of EVSI in real-world scenarios using nested Monte Carlo (MC) sampling [6]. This computational burden is ~~further increased when one aims to compute EVSI for multiple trial designs in order to determine the optimal~~ increased further when we aim to identify the optimal research study (i.e., ~~the study~~ with the highest ENBS) ~~research study~~ and must compute EVSI for multiple trial designs [7, 8]. High performance computing ~~resources~~ can be used to overcome some of these barriers, but ~~often at the expense of an increased requirement for~~ this requires additional programming skills and ~~an increase in~~ increases the complexity of the analysis.

Several methods have been developed to ~~unlock the potential of EVSI as a tool~~ overcome these computational barriers and, thus, use EVSI for research prioritization and trial design optimization ~~by overcoming these computational barriers. [9–19].~~ [9–20]. However, as many of these methods have been developed concurrently, they have not been ~~directly compared~~ compared, making it challenging to select an appropriate method. Additionally, ~~the initial publications for these EVSI estimate methods typically evaluate these methods~~ these EVSI estimation methods have mostly been evaluated using health economic models and trial designs chosen for computational convenience rather than to reflect real-world decision making. Thus, the accuracy of the majority of these methods has not been assessed in, for example, multi-state models combined with survival or ~~quality of life outcomes. A~~ quality-of-life outcomes, a setting that would occur frequently in real-life health economic modelling.

We aim, therefore, to address these gaps by comparing the relative performance of EVSI calculation methods across three realistic health economic models and trial designs to gain a greater understanding of their behaviour in practice. We will evaluate the accuracy of the EVSI estimation methods across ~~these~~ our three examples and the computational time required to obtain these estimates. These examples ~~have several key~~ exhibit features that reflect real-world trial design and may ~~make it challenging to estimate EVSI~~ influence the behaviour of these EVSI estimation methods 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. ~~This comparison can aid researchers in selecting between these methods while also ensuring they can be used to estimate EVSI accurately in practice.~~

Due to the complexity of the three case studies, our comparison is restricted to four recent calculation methods ~~developed by (in chronological order)~~ the Regression Based (RB) method developed by Strong et al. [12], the Importance Sampling (IS) method developed by Menzies [13], the Gaussian Approximation (GA) method developed by Jalal and Alarid-Escudero [15] (extending a method proposed in Jalal et al. [14]), and the Moment Matching (MM) method developed by Heath et al. [17–19]. These methods are ~~all~~ based on different approaches and assumptions, but they all provide ~~estimation techniques for approximating EVSI that, in comparison with~~ EVSI estimates with a smaller computational burden compared to the nested MC sampling methods ~~, are less computationally demanding~~ whilst retaining accuracy.

~~Alternative EVSI methods either place restrictions on the structure of the underlying health economic model, the number of evidence sources and/or the study design [9–11]. 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. EVSI estimation methods based on minimal modelling, where a comprehensive clinical trial is available to inform EVSI estimation, have also been proposed [20]. These methods are competitive~~

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 some measure of benefit (which is usually net monetary benefit or net health benefit [21]). We characterize a health economic model as a function that takes the vector of parameters $\boldsymbol{\theta}$ as an input, 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 $p(\boldsymbol{\theta})$. 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 $\boldsymbol{\theta}$. Uncertainty about $\boldsymbol{\theta}$ induces uncertainty about the net benefit for each treatment $t = 1, \ldots, T$. We denote the net benefit for treatment t given parameters $\boldsymbol{\theta}$ as $\mathrm{NB}_t^{\boldsymbol{\theta}}$. 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 we are interested in collecting additional information about a subset of the model parameters $\boldsymbol{\phi}$. In this ~~settings~~ setting, $\boldsymbol{\theta}$ is split into two sets of parameters $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\psi})$, where $\boldsymbol{\psi}$ are all the remaining model parameters not in $\boldsymbol{\phi}$. 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 $\boldsymbol{\phi}$ (assuming risk neutrality) is equal to the Expected Value of Partial Perfect Information (EVPPI) [22–24]. This is given by

$$\mathrm{EVPPI} = \mathrm{E}_{\boldsymbol{\phi}} \left[\max_t \mathrm{E}_{\boldsymbol{\theta} | \boldsymbol{\phi}} \left[\mathrm{NB}_t^{\boldsymbol{\theta}} \right] \right] - \max_t \mathrm{E}_{\boldsymbol{\theta}} \left[\mathrm{NB}_t^{\boldsymbol{\theta}} \right]. \tag{4}$$

In this paper, EVSI is defined as the value of collecting additional data, denoted $\boldsymbol{X}$, to inform $\boldsymbol{\phi}$, where $\boldsymbol{\phi}$ could be equal to $\boldsymbol{\theta}$ if the additional data updates all the underlying model parameters. By assuming that $\boldsymbol{X}$ directly updates $\boldsymbol{\phi}$ only, we implicitly state that $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{\psi} \mid \boldsymbol{\phi}$. This implies that EVSI is bounded above by the EVPPI for $\boldsymbol{\phi}$. If these data $\boldsymbol{X}$ had been collected and observed to have a value $\boldsymbol{x}$, they would be combined with the current evidence to generate an updated distribution for $\boldsymbol{\phi}$, $p(\boldsymbol{\phi} \mid \boldsymbol{x})$. 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 $\boldsymbol{x}$ is the treatment associated with the maximum expected net benefit based on the updated knowledge about the relevant parameters $\boldsymbol{\phi}$. If the optimal intervention changes, compared to the current decision, then the information in $\boldsymbol{x}$ 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

$$\mathrm{EVSI} = \mathrm{E}_{\boldsymbol{X}} \left[\max_t \mathrm{E}_{\boldsymbol{\theta} | \boldsymbol{X}} \left[\mathrm{NB}_t^{\boldsymbol{\theta}} \right] \right] - \max_t \mathrm{E}_{\boldsymbol{\theta}} \left[\mathrm{NB}_t^{\boldsymbol{\theta}} \right], \tag{5}$$

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

Calculation Methods for EVSI

It is rarely possible to compute EVSI analytically as the net benefit is often a complex function of $\boldsymbol{\theta}$. Additionally, it is challenging to compute the expectation of a maximum analytically as required in the first term of equation (5). Therefore, a range of methods have been developed to approximate EVSI.

Nested Monte Carlo Computations for EVSI

The simplest approximation method [6] computes all the expectations in equation (5) using MC simulation. The second term can be computed by simulating $s = 1, \ldots, S$ parameter values, $\boldsymbol{\theta}_s$, from $p(\boldsymbol{\theta})$. The simulated values are used as inputs to a health economic model to obtain S simulations of the net benefit for each intervention, denoted $\mathrm{NB}_t^{\boldsymbol{\theta}_s}$. Note that this process is required to perform a

"probabilistic analysis" (PA) [25], used to assess the impact of parametric uncertainty on the decision uncertainty, which is mandatory in various jurisdictions [26–28]. The average of $\mathrm{NB}_t^{\boldsymbol{\theta}_1}, \ldots, \mathrm{NB}_t^{\boldsymbol{\theta}_S}$ for each intervention can be computed and $\max_t \mathrm{E}_{\boldsymbol{\theta}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]$ is estimated by the maximum of these means.

The first term in equation (5) is more complex to compute by simulation. Firstly, S datasets $\boldsymbol{X}_s$ must be generated conditional on the simulated $\boldsymbol{\theta}_s$ from the assumed sampling distribution $p(\boldsymbol{X} \mid \boldsymbol{\theta}_s)$. For each $\boldsymbol{X}_s$, we simulate R values from the updated distribution of the model parameters $p(\boldsymbol{\theta} \mid \boldsymbol{X}_s)$. 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 $\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right]$ for $t = 1, \ldots, T$. The maximum of these simulated means is then selected for each $\boldsymbol{X}_s$. 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 net benefit associated with each intervention t using a smaller simulation burden. We denote the expectation of the net benefit, conditional on data $\boldsymbol{X}$, as

$$\mu_t^{\boldsymbol{X}} = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right].$$

In a similar manner, we also denote the expectation of the net benefit, conditional on some value of the parameters of interest ϕ, as

$$\mu_t^{\boldsymbol{\phi}} = \mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{\phi}}\left[\mathrm{NB}_t^{\boldsymbol{\theta}}\right].$$

Finally, we can increase the numerical stability of the following approximation methods by working in terms of the incremental net benefit or loss, defined, without loss of generality, as $\mathrm{INB}_t^{\boldsymbol{\theta}} = \mathrm{NB}_t^{\boldsymbol{\theta}} - \mathrm{NB}_1^{\boldsymbol{\theta}}$ for $t = 2, \ldots, T$. This is because we must estimate $\mu^{\boldsymbol{X}}$ for each of the $T - 1$ incremental net benefit options rather than the T net benefits. EVSI can then be estimated by setting $\mu_1^{\boldsymbol{X}} = 0$ and calculating

$$\widehat{\mathrm{EVSI}} = \frac{1}{N}\frac{1}{S} \sum_{s=1}^{S} \max_t \mu_t^{\boldsymbol{X}_s} - \max_t \frac{1}{N}\frac{1}{S} \sum_{s=1}^{S} \mu_t^{\boldsymbol{X}_s}, \tag{6}$$

where $\mu_t^{\boldsymbol{X}_s}$ is the estimated posterior expectation for the dataset $\boldsymbol{X}_s$, $s = 1, \ldots, S$.

~~Strong et al. (2015)~~ Regression Based Method

The ~~Strong et al. method~~ RB method [12] estimates EVSI by fitting $T - 1$ regression models. For each $t = 2, \ldots, T$, the simulated values of the incremental net benefit are the 'response' variable and a low-dimensional summary of the simulated dataset $\boldsymbol{X}$ is the '~~independent' or '~~predictor' variable(s) [12]. This low-dimensional summary for $\boldsymbol{X}$ should reflect how the data would be summarized if the study were to go ahead and must be computed for each simulated dataset $\boldsymbol{X}_s$. The fitted values from this regression model are then used to estimate $\mu_t^{\boldsymbol{X}}$ ~~is estimated by the fitted values from this regression model~~. EVSI is then estimated directly from these estimates of $\mu_t^{\boldsymbol{X}}$ using equation (6).

~~Menzies (2016)~~ Importance Sampling Method

~~Menzies [13] presents two EVSI estimation methods~~ Initially, two IS methods were proposed to estimate EVSI [13], the most accurate of ~~which~~ these estimates $\mu_t^{\boldsymbol{X}}$ by reweighting simulations of $\mu_t^{\boldsymbol{\phi}}$. This reweighting is based on the *likelihood* of observing a simulated dataset ~~$\boldsymbol{X}$~~ $\boldsymbol{X}_s$ conditional on different values for $\boldsymbol{\phi}$. The term likelihood is used in the statistical sense and is equal to $p(\boldsymbol{X} \mid \boldsymbol{\phi})$.

This method simulates S future datasets $\boldsymbol{X}_s$ from $p(\boldsymbol{X} \mid \boldsymbol{\phi}_s)$. The likelihood for *every* simulated vector for $\boldsymbol{\phi}$ is then calculated conditional on $\boldsymbol{X}_s$. For the sample $\boldsymbol{X}_s$, $\mu_t^{\boldsymbol{X}_s}$ is estimated as the weighted average of $\mu_t^{\boldsymbol{\phi}}$, weighted by the likelihood of the dataset $\boldsymbol{X}_s$. Thus, the ~~Menzies method can be seen as IS method is~~ an example of importance sampling [29, 30] and performs $(T - 1)S^2$ likelihood calculations for each EVSI estimate. EVSI is estimated from equation (6) based on the estimate of $\mu_t^{\boldsymbol{X}_s}$ for each future sample.

Gaussian Approximation Methods

The GA method [14, 15] fits a "linear" metamodel[3], a secondary model that captures the relationship between the simulated incremental net benefit values, as the response variable, and simulations for ϕ, as the predictor variables. Each term of the linear metamodel is then rescaled based on a Gaussian-Gaussian Bayesian updating approach to estimate its 'posterior' expectation across different future datasets $\boldsymbol{X}$. These estimated distributions are then recombined using the coefficients of the linear metamodel to estimate $\mu_t^{\boldsymbol{X}}$ and compute EVSI.

For a proposed future data collection strategy of size N, the rescaling factor for each term of the linear metamodel is equal to

$$\frac{N}{N + N_0},$$

where N_0 is known as the prior effective sample size [31]. Essentially, N_0 represents the number of independent observations who would be required to generate the amount of evidence in the prior. In some prior-likelihood pairs, N_0 can be obtained analytically. In other settings, Jalal *et al.* suggest two estimation methods for N_0. Firstly, if the data $\boldsymbol{X}$ can be summarized using a summary statistic $W(\boldsymbol{X})$, then N_0 can be computed as a function of the variance of $W(\boldsymbol{X})$. Secondly, if a suitable statistic cannot be derived, then nested posterior sampling can be used to estimate N_0. In this method, S future datasets $\boldsymbol{X}_s$, $s = 1, \ldots, S$ are simulated. Each of these samples is used to update the information about the model parameters $p(\boldsymbol{\theta} \mid \boldsymbol{X}_s)$, typically using R simulations and computing the mean for ϕ. The *variance* of the mean for ϕ, across different samples $\boldsymbol{X}_s$, is then used to estimate N_0. Computationally, this nested sampling method to compute N_0 is relatively 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.

Moment Matching Method

The MM method [18, 19] combines the simulations μ_t^{ϕ} 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 [19]. Thus, EVSI is estimated with $Q \times R$ health economic model runs.

The MM method uses nested MC sampling to estimate the *variance* of the incremental net benefit for different future datasets. These estimated variances rescale simulations of μ_t^{ϕ} for $t = 2, \ldots, T$ to approximate simulations of $\mu_t^{\boldsymbol{X}}$ which can be used to estimate EVSI using equation (6). The MM method only requires a single nested simulation procedure to estimate EVSI across different sample sizes of the future trial [19].

Of note, this comparison has been restricted to these four methods as the alternative EVSI methods typically place restrictions on the structure of the underlying health economic model, the number of evidence sources and/or the study design [9–11, 11, 20]. These restrictions limit the applicability of these methods. Some of these methods make assumptions about the distribution of the parameters and the study data to ensure that the prior and posterior model parameter distributions take the same form (conjugacy). This allows for computationally efficient EVSI estimation. The minimal modelling approach [20] assumes that a comprehensive clinical trial is available to inform EVSI estimation and, thus, restricts the required data sources. These methods are competitive in terms of computational time and accuracy compared to the methods presented in this review but are restricted to the settings in which they are relevant which limits their general purpose application.

Case Studies

These EVSI methods are applied to three case studies designed to explore different trial designs and health economic models. These designs and models were chosen to assess the accuracy of these EVSI estimation methods in settings that are reflective of real-world decision making. The first case study is a stylized Chemotherapy side effects example used to

---

[3]A linear metamodel is required for this method. However, non-linear functions of ϕ can be defined and combined linearly to account for flexible relationships between the incremental net benefit and the parameters ϕ.

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~~ Model for Chemotherapy ~~Treatment~~Side Effects

This model was presented in Heath and Baio [18] 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 π_0 and ρ 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, γ_1 and γ_2 denote the constant probability of requiring hospital care and dying, respectively, and λ_1 and λ_2 denote the constant probability of recovery given that an individual ~~remains at home or enters~~ is not or is admitted into hospital, respectively. The cycle length is 1 day~~and~~, the time horizon is 15 days, and we assume that only hospitalised patients will die. 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. PA distributions for the model parameters are informed using previous data or defined using expert opinion with all distributional assumptions given in the supplementary material.
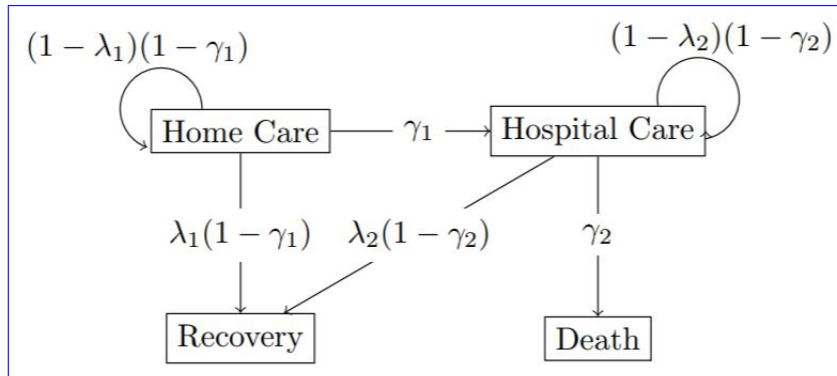

Figure 5: A four state Markov model used to model the health economic impact of adverse events from a chemotherapy treatment.

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 with 150 patients per arm.

To define the sampling distribution for the six outcomes, we model the number of patients who experience adverse events using binomial distributions conditional on π_0 and ρ;

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

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

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

Finally, recovery time for patients who experience adverse events but recover whilst remaining at home is modelled with an exponential distribution conditional on the transition probability λ_1,

$$T_{HC}^i \sim Exponential(\eta_1)$$

for each individual who remains at home whilst recovering from adverse events $i = 1, \ldots, X_{AE_0} + X_{AE_1} - X_{Hosp}$ and $\eta_1 = -\log(\lambda_1)$. The recovery time for every patient ($j = 1, \ldots, X_{Hosp} - X_{Death}$) who recovers in hospital is modelled as and exponential distribution conditional on λ_2

$$T_H^j \sim Exponential(\eta_2)$$

with $\eta_2 = -\log(\lambda_2)$. Exponential distributions were used to model the recovery time as we assumed a constant transition probability in the Markov model.

Case Study 2: A Model for Chronic Pain

This example uses a cost-effectiveness model developed by Sullivan et al. [32], and extended in Heath *et al.* [19], 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. Patients initially receive treatment for chronic pain and can either experience adverse events or not. Patients may then withdraw from this initial treatment due to adverse events or lack of efficacy. Following this, they can be offered an alternative therapy or withdraw completely from treatment. Patients may experience adverse events from this second-line treatment and can either withdraw from this treatment due to these adverse events or lack of efficacy. If patients withdraw from this second-line treatment, they can receive further treatment or discontinue, both considered absorbing states as the model does not include a death state.

Initially, patients can either be offered morphine or an innovative treatment, and our model evaluates the cost-effectiveness of the innovative treatment. If patients require a second-line treatment, they are offered oxycodone. Thus, the only difference between the two treatment 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 [32] where the distributions for the PA are gamma for costs and beta for probabilities and utilities. In the original publication of this model, the means of these distributions were informed by relevant studies identified following a literature review and the standard error of these mean estimates was taken as 10% of the underlying mean estimate. This paper computed the per-person lifetime EVSI, 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 for patients who withdraw from the first-line treatment due to lack of efficacy. The individual level variability in these two QoL weights is modelled, for simplicity, as independent beta sampling distributions although the assumption of independence may be invalid [33]. The population level mean QoL weight, i.e., the mean of the sampling distribution for the QoL weights, is defined as the value of those two health states in the Markov Model. The standard deviations of the individual level sampling 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 [34][4]. 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 [35]. 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 Screening

This example uses a health economic model developed by Alarid-Escudero et al. [36] to evaluate a screening strategy for colorectal cancer (CRC) and pre-cancerous lesions known as adenomas. The model is based on a nine-state Markov model with age-dependent transition intensities which govern the

---

[4]This sampling distribution for the data causes some minor issues for the Gibbs sampling procedure used in the JAGS program for Bayesian updating.

onset of adenomas (pre-cancerous growths) and the risk of all-cause mortality. The onset of adenomas is ~~modeled~~ modelled using a Weibull hazard conditional on age

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

where λ_1 and g are the shape and scale parameters of the Weibull distribution and a is the age of the patient. Model parameters are chosen to reflect the prevalence of adenomas seen in the literature. To determine the PA distribution for the model parameters g and λ_1, the level of uncertainty in the observed prevalence was characterized and the parameters were re-estimated across the distribution of prevalence.

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. When the model is initiated, some members of the general population have undiagnosed adenomas and early stage CRC.

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 [37], were used to generate study subjects with an age distribution representative of the general population, with study ~~enrollment~~ enrolment 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 $p(a_i)$

$$X_i \sim Ber\left(p\left(a_i\right)\right).$$

~~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~~ It is challenging to compare the four efficient 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. Thus, Table 3 outlines the simulation choices that were made for the case studies ~~. These choices were made to achieve EVSI estimates~~ in this paper. We chose these assumptions to reflect realistic choices that would be made by modellers calculating EVSI in practice. They all aim to estimate EVSI 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~~ Based on these choices, we compared the speed and accuracy ~~achievable by~~ of each method, and identified their relative advantages and challenges in practice.

~~These choices impact our results as larger sample sizes would lead to greater computational time and accuracy for all methods. Depending on the method and the underlying case study, the impact of sample size will change. For example, more simulations are needed for the Strong et al. and Jalal et al. methods if there are a large number of outcomes in the proposed trial, increasing the complexity of the regression model or the metamodel, and for the Heath et al. method if the underlying EVSI is small compared to the overall value of resolving parametric uncertainty. Nonetheless, we believe that the choices highlighted in Table 3 represent a fair comparison of these methods in practice~~ The Chemotherapy side effects and Chronic pain models had low computational cost and, thus, we selected a large PA simulation size S as VoI methods require an accurate representation of the decision uncertainty [38]. The CRC screening model had greater computational cost and thus, we used a smaller simulation size to reflect choices that would be made in practice. In all case studies, we used all the PA simulations to calculate EVPPI and estimate μ_i^{ϕ}.

For the MC nested simulation, we used all the PA simulations for the outer simulation loop. To accurately characterise the PA distribution for the inner loop, we use the same size as the outer loop. This gives a very high simulation burden but leads to accurate results for EVSI that can be compared with the novel computation methods.

Simulation Choices	Case Study		
	Chemotherapy side effects (1)	Chronic ~~Pain~~ pain (2)	CRC screening (3)
Initial PA size (S)	100,000	100,000	5,000
Number of μ_t^ϕ simulations from EVPPI calculation	100,000	100,000	5,000
Nested simulation outer loop size (S)	100,000	100,000	NA
Nested simulation inner loop size (R)	100,000	100,000	NA
~~Strong et al.~~ RB simulation size	100,000	100,000	5,000
~~Menzies~~ IS simulation size	20,000	5,000	2,500
~~Jalal et al.~~ GA N_0 computation method	nested posterior sampling	nested posterior sampling	nested posterior sampling
~~Jalal et al.~~ GA N_0 estimation outer loop size	1,000	1,000	5,000
~~Jalal et al.~~ GA N_0 estimation inner loop size	10,000	10,000	5,000
~~Jalal et al.~~ GA N_0 estimation future sample size	30	40	40
~~Heath et al.~~ MM outer loop size (Q)	50	50	50
~~Heath et al.~~ MM inner loop size (R)	10,000	10,000	5,000

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

~~Note that the simulation size for the Menzies method is typically smaller than for the other methods as the computational cost increases with S^2. This makes it computationally challenging for~~

For all case studies, the RB method uses all the PA simulations as regression modelling requires larger simulation sizes ~~.~~ to accurately estimate EVSI. The computational cost of using all the simulations is minimal. In contrast, we used a smaller simulation size for the IS method as it has a higher computational cost compared to the other methods that increases proportional to S^2 and we wanted to ensure a comparable analysis in terms of computation time.

For the ~~Strong et al. method requires larger simulation sizes to accurately estimate EVSI. For all three case studies the Jalal et al. methods requires~~ GA method, we used nested posterior sampling to estimate the prior effective sample size ~~. To decrease the computational time of this estimation, we note that~~ for all three case studies. N_0 is estimated using a proposed future sample with a given sample size but only 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~~ we can reduce the computational cost of estimating N_0 ~~with a small~~ by using a small sample size for the proposed sample $\boldsymbol{X}$. However, as the estimation of N_0 also relies on a Gaussian approximation~~so~~, the sample size of $\boldsymbol{X}$ should be sufficiently large to assume normality. Thus, ~~the table above (Jalal et al. future sample size) highlights the sample size of $\boldsymbol{X}$ used in the nested posterior sampling to estimate N_0 that balances accuracy and computational speed. Finally, the Heath et al. loop sizes were chosen based on Heath and Baio [19], which highlights that $Q = 50$~~

we selected future sample sizes for the GA updating of around 30 (or 40 to adjust for missingness) to ensure the assumption of normality holds but to decrease computational complexity. The simulation sizes for the nested simulation were chosen to balance accuracy and computation time. Specifically, we chose more simulations for the CRC screening model as the posterior of the parameters for a Weibull distribution is highly correlated and therefore challenging to estimate using Gibbs sampling [39].

Finally, Heath and Baio determined that $Q = 50$ is sufficient to estimate EVSI using MM method, provided the inner loop size R is sufficient to capture the posterior for the model parameters ~~is adequately captured.~~ [19]. Thus, the inner loop sizes were chosen relatively large to capture the posterior distribution of the model parameters, with the CRC screening model simulation size reduced due to the computational complexity of that example.

~~For the first two case studies, we~~ In general, larger PA simulation sizes lead to increased computational time and accuracy for all methods. The improvements associated with larger sample sizes change depending on the method and the case study. For example, the RB and GA methods require a greater number of PA simulations if there are a large number of outcomes in the proposed trial as the regression model or the metamodel increases in complexity. Elsewhere, the MM method requires more PA simulations if the underlying EVSI is small compared to the overall value of resolving parametric uncertainty. Nonetheless, we believe that the choices highlighted in Table 3 represent a fair comparison of these methods that supports the use of these methods based on modelling choices that could be implemented in practice.

To characterise uncertainty in the EVSI estimation procedure for each method, we computed a standard error for the EVSI estimates by recomputing the EVSI 200 times, ~~each time with~~ based on the same PA ~~samples, so that this standard error reflects uncertainty arising from any simulation involved in the EVSI estimation procedure itself.~~

~~To obtain the~~ simulations for the first two case studies. The computational time for the four recent approximation methods ~~, computations were~~ is presented based on computations 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~~ computation time for the nested MC computation is based on the total computation time required across 32 cores on the Hospital for Sick Children's High Performance Computing environment []. Code to undertake the computations in this paper is available from GitHub at `https://github.com/convoigroup/EVSI-in-practice`.

Results

Case Study 1: Chemotherapy Side Effects

Figure 6 displays the 95% central intervals for the four ~~faster~~ fast 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"~~ RB and MM methods contain the 'true' value, represented by the nested MC EVSI. ~~At the same time, the Heath et al.~~ In this example, however, the MM estimate is associated with substantial variability compared to the other methods as the EVSI is smaller compared to the overall EVPI.

Implementing the ~~Strong et al. and Jalal et al.~~ RB and GA 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 7 shows that the 95% central intervals for the ~~Heath et al. and the Menzies~~ MM and the IS methods contain the nested MC estimate ~~, which we assume to be accurate given the large simulation size,~~ for all sample sizes. In addition to this, EVSI calculated by all four methods increase as the sample size increases, representing that the more information is collected, the greater ~~value the value~~ its value. EVSI estimates should also remain below EVPPI (marked as a dashed line on Figure 7), which represents an upper limit on EVSI. The EVSI estimates are also relatively close to the nested MC estimate. The ~~Strong et al.~~ RB method produced the shortest 95% central intervals while the three alternatives are
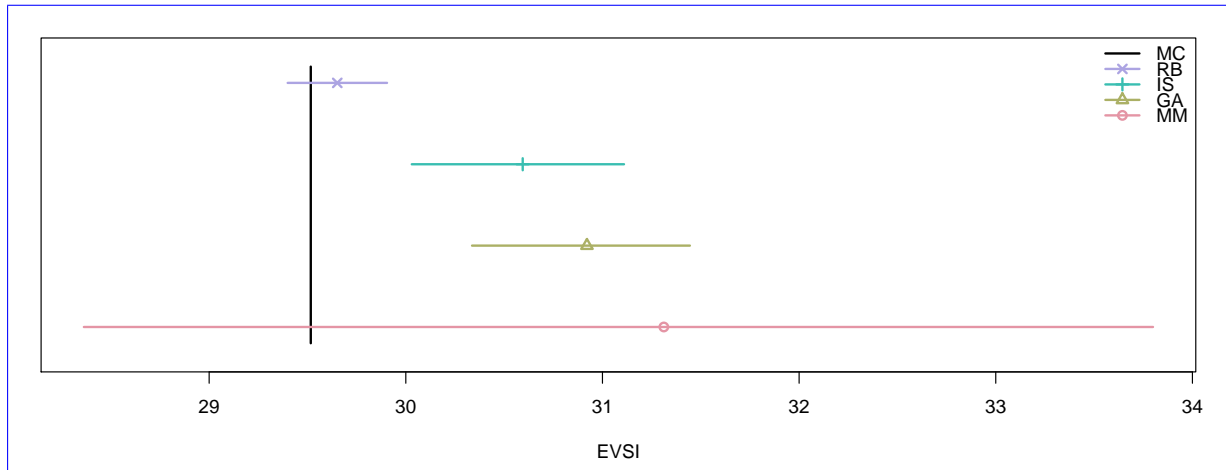
Figure 6: The mean per-person EVSI estimates, across 200 simulated estimation procedures, for the five methods under consideration for the ~~chemotherapy~~ Chemotherapy side effects example with a future sample size of 150 and willingness-to-pay of £30,000. The 95% central intervals from these 200 simulations are shown as horizontal lines and the gold standard MC estimator is shown as a vertical line.

relatively comparable. Note that the ~~Menzies~~ IS estimate is based on a smaller PA simulation size but still offers similar variability compared to the other methods.

In this example, the summary statistic used for the ~~Strong et al.~~ RB 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 [40]. Summarizing X using the arithmetic mean and variance gives incorrect EVSI estimates for this case study as these statistics are not sufficient. Low-dimensional sufficient statistics should be used, when available, to estimate EVSI if the proposed analysis following the trial will use Bayesian methods to incorporate the additional information into the evidence base of the health economic model.

Case Study 3: Colorectal Cancer Screening

Figure 8 demonstrates EVSI estimates (y-axis) for the CRC screening model across the considered sample sizes (x-axis, on the log-scale but marked on the natural scale). We can see that the four EVSI calculation methods give a broad consensus. 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."~~ 'correct.'

~~For~~ Of note, for a sample size of 1,500, the ~~Menzies~~ IS EVSI estimate is incorrect. This is because the likelihood tends to 0 for large sample sizes making the ~~weighted samples difficult~~ weights in the weighted sum challenging to approximate. Furthermore, the ~~Menzies~~ IS method slightly over-estimates the EVSI for sample sizes between 500 and ~~1000.~~ 1,000. This is because we only use a subset of the PA 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 PA simulations.

Computational Time

Table 4 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 screening 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.~~ MM method has the lowest computation time as the underlying health economic model is fast. The ~~Heath et al.~~ MM method also estimates EVSI across multiple sample sizes simultaneously which improves the computational time for the Chronic ~~Pain~~ pain example compared to the ~~Strong et al., and Menzies~~ RB, and IS methods. For these two examples, the computation time required to fit an accurate regression model is relatively high, increasing
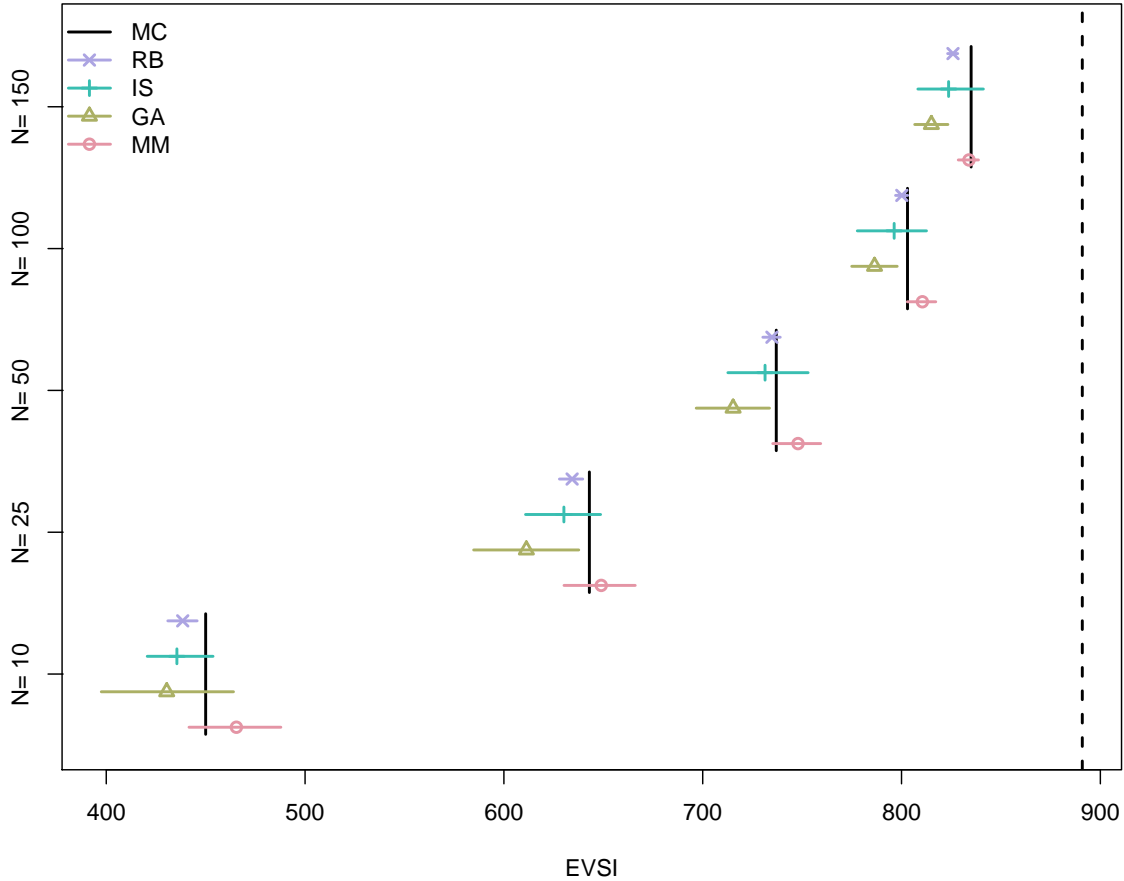
Figure 7: The mean EVSI estimates, across 200 simulated estimation procedures, for the five methods under consideration for the ~~chronic~~ Chronic pain example. EVSI was calculated across 5 different sample sizes for the future trial. The 95% central intervals from these 200 simulations are shown as horizontal lines and the gold standard MC estimator is shown as a vertical line.

the computation time for the ~~Strong *et al.*~~ RB method. The ~~Jalal *et al.*~~ GA 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.*~~ GA method would offer computational savings on the ~~Strong *et al.*, and Menzies~~ RB and IS methods. For the Chemotherapy side effects example, the ~~Menzies~~ IS 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 PA simulations are used, the computation time is greater than 2 hours. For the Chronic ~~Pain~~ pain example, the ~~Menzies~~ IS method is noticeably slower as the computation time for the likelihood increases when the proposed sample size of $\boldsymbol{X}$ is larger.

Case Study	Computational Time (mins)				
	Nested MC	RB	IS	GA	MM
1: Chemotherapy side effects	37646	5.35	4.56	8.20	2.01
2: Chronic pain	223200	12.05	86	22.27	2.46
3: CRC screening	*	27.24	91	7.17	492

Table 4: 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.*~~ GA 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.*~~ RB method, $\boldsymbol{X}$ is summarized by finding the maximum
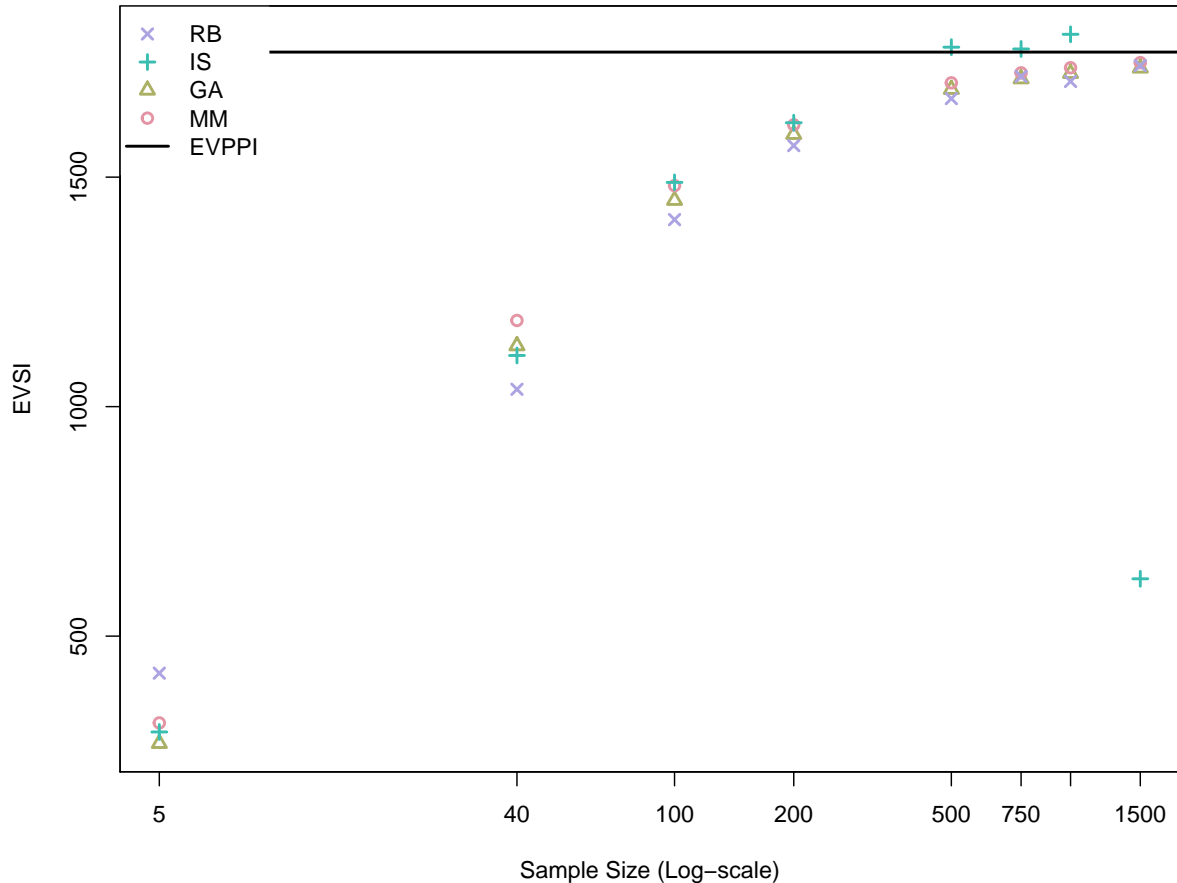
Figure 8: EVSI estimates for the four methods under consideration for the CRC screening model. EVSI is calculated for 9 different sample sizes for the future trial and is plotted across sample size. The sample size is plotted on the log scale with the sample sizes marked on the natural scale. The EVPPI, computed using the Strong *et al.* EVPPI computation method [41], is included as a black line on this Figure.

likelihood estimates (MLE) for g and λ_1 that must be estimated, using relatively slow computational optimization procedures, for each sample $\boldsymbol{X}_s$, $s = 1, \ldots, S$ and sample size. Thus, estimating the summary statistics is slow in this case study. The ~~Heath *et al.*~~ MM method is more computationally expensive as the underlying probabilistic sensitivity analysis for the CRC screening health economic model is expensive and must be rerun $Q \times S = 250,000$ times to compute EVSI. The computational time of the ~~Menzies~~ IS method is similar to the previous case studies.

Discussion

The paper uses three case studies to ~~assess~~ compare the accuracy and computational time for 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~~ as no previous head-to-head comparison had ~~previously been undertaken~~ been done. Additionally, ~~these methods have typically been assessed~~ we assessed the performance of these methods 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~~ that were designed to cover a number of different trial designs, interventions and health economic model structures. Thus, we assessed the comparative performance of these efficient EVSI estimation methods across a number of scenarios that may make the EVSI estimation more challenging.

In general, the EVSI estimates ~~were~~ are accurate when the underlying assumptions for the respective methods were met ~~, highlighting the importance of checking these assumptions. It should be noted,~~

~~however, that appropriately assessing accuracy is challenging because differences in EVSI estimates could lead to alternative future research recommendations, even when the difference is small. The appropriate level of accuracy for EVSI estimates will, therefore, depend on the decision problem. Estimates should be accurate enough to ensure that the future research recommendation would not be expected to change if the estimates are made more accurate.~~

~~In general, the~~ and can, thus, be used with confidence in practice. 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. ~~This analysis has highlighted that the Menzies~~ Although, the IS method is generally more computationally intensive than the other methods.

~~Additionally, this analysis has highlighted that there is not a single systematically superior~~ In this analysis, we have not identified an efficient EVSI estimation method that is systematically superior to the alternatives. Specifically, the "optimal" estimation method ~~trading~~ that trades off accuracy, precision, computational time and ease of implementation will change depending on the health economic model structure, proposed trial design and analyst expertise ~~. Thus, these methods should also be compared on their ease of implementation.~~ [42].

~~While future research is required to fully articulate the relative strengths and limitations of these methods for different health economic models and trial designs~~ Nonetheless, the analysis in this paper has emphasized some distinctions between the methods. Firstly, the ~~Strong et al.~~ RB method is accurate and efficient, provided the analyst can correctly summarize the trial data and fit a regression model, which ~~in some settings~~ can be challenging either statistically or computationally. The ~~Menzies~~ IS method is accurate but computationally expensive for large PA simulation sizes. The ~~Jalal et al.~~ GA method is efficient when estimating EVSI across sample ~~size~~ sizes but often requires nested posterior sampling when considering realistic data collection exercises and ~~also~~ relies on an accurate "linear" ~~model~~ metamodel. Finally, the ~~Heath et al.~~ MM method is accurate and efficient when the health economic model has a low computation time but becomes more unfeasible as the model run time increases. The ~~Jalal et al. and Heath et al.~~ GA and MM methods require expertise in Bayesian methods/thinking. The ~~Menzies~~ IS method requires the repeated evaluation of the likelihood function, and hence relies on this function being both known and computationally tractable.

Beyond these remarks, future research is required to fully articulate the relative strengths and limitations of these methods for different health economic models and trial designs and to support the widespread implementation of these methods.

A   Inputs for the Chemotherapy Side Effects Model

Model Input	Distribution	1st Prior Parameter	2nd Prior Parameter	Previous Data
π_0 - 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	No
Γ_1 - Probability of hospitalization	Beta	1	1	Number of hospitalizations
Γ_2 - Probability of death	Beta	1	1	Number of deaths
γ_1 - Daily transition probability to hospital	$\frac{\Gamma_1}{15}$	-	-	-
γ_2 - Daily probability of death	$\frac{\Gamma_2}{15}$	-	-	-
λ_1 - Daily probability of recovery from home care	Beta	5.12	6.26	No
λ_2 - Daily probability of recovery from hospital	Beta	3.63	6.74	No
Cost of death	LogNormal	8.33	0.13	No
Cost of home care	LogNormal	7.74	0.039	No
Cost of hospitalization	LogNormal	8.77	0.15	No
QoL weight for home care	Beta	5.75	5.75	No
QoL weight for hospitalization	Beta	0.87	3.47	No

Table 5: 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.