Bayesian computations for Value of Information measures using Gaussian processes, INLA and Moment Matching

Anna Louise Heath

Thesis for Submission at University College London
Research Degree: Statistical Science

DECLARATION

I, Anna Heath, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Date: 23rd March 2018

To my Grandpa David

ACKNOWLEDGEMENTS

No (wo)man is an island and clearly this thesis is the product of hard work from a number of different people, so bear with me while I get the “thank yous” out of the way!

First and foremost, I would like to thank my supervisors, Gianluca and Ioanna, for their hard-work, dedication and endless support of me throughout this process. To have to opportunity to work with two such intelligent, kind and passionate people has been an absolute pleasure and I couldn’t have done it without them. Every meeting, paper edit, e-mail and notation discussion(!) has gone into making this thesis. Some students don’t get to work with one supervisor who show the level of interest you’ve had in this work so I’ve been unbelievably lucky to have two.

Secondly, I’d like to thank all the anonymous (and not-so-anonymous) reviewers who have taken the time to comment on my work and infinitely improved the quality. In addition to this, my thanks go (in no particular order) to Nicky Welton and Mark Strong for their discussions on all aspects of Value of Information, to Will Sullivan for the Chronic Pain model, to Nick Menzies for his MC simulations, to Alice Davis for her help on the SPDE-INLA methodology and to Rachael Hunter for her work on the EVSI package.

On a more personal level, I’d like to thank all my colleagues in Room 103 and beyond for their companionship throughout these 3 and a bit years. I’d also like to thank all the support staff in the department for their help. My thanks also go to my family for their help and support, particularly for convincing me that everyone’s thesis goes badly at times and proudly telling all about the book! I want to thank my Grandpa who sadly passed away before this thesis was completed but was dedicated to reading my papers and tracking their citation count! Last but not least, I want to thank my partner Sam for all his support throughout this process, he knows more about Value of Information than anyone would wish; I couldn’t have done this without him!

ABSTRACT

Value of Information measures quantify the economic benefit of obtaining additional information about the underlying model parameters of a health economic model. Theoretically, these measures can be used to understand the impact of model uncertainty on health economic decision making. Specifically, the Expected Value of Partial Perfect Information (EVPPI) can be used to determine which model parameters are driving decision uncertainty. This is useful as a tool to perform sensitivity analysis to model assumptions and to determine where future research should be targeted to reduce model uncertainty.

Even more importantly, the Value of Information measure known as the Expected Value of Sample Information (EVSI) quantifies the economic value of undertaking a proposed scheme of research. This has clear applications in research prioritisation and trial design, where economically valuable studies should be funded. Despite these useful properties, these two measures have rarely been used in practice due to the large computational burden associated with estimating them in practical scenarios. Therefore, this thesis develops novel methodology to allow these two measures to be calculated in practice.

For the EVPPI, the method is based on non-parametric regression using the fast Bayesian computation method INLA (Integrated Nested Laplace Approximations). This novel calculation method is fast, especially for high dimensional problems, greatly reducing the computational time for calculating the EVPPI in many practical settings. For the EVSI, the approximation is based on Moment Matching and using properties of the distribution of the *preposterior* mean. An extension to this method also uses Bayesian non-linear regression to calculate the EVSI quickly across different trial designs. All these methods have been developed and implemented in `R` packages to aid implementation by practitioners and allow Value of Information measures to inform both health economic evaluations and trial design.

OUTPUTS

Publications

- **Heath A.**, Manolopoulou, I., Baio, G. “Efficient Monte Carlo Estimation of the Expected Value of Sample Information using Moment Matching” *Medical Decision Making*, 2017.
- Baio, G., Berardi, A., **Heath, A.** “Bayesian Cost-Effectiveness Analysis with the R package BCEA" *Springer*, 2017.
- **Heath A.**, Manolopoulou, I., Baio, G. “A Review of Methods for the Analysis of the Expected Value of Information" *Medical Decision Making*, 2016.
- Baio, G., **Heath A.**“When Simple becomes Complicated: Why Excel should lose its place at the top table". *Global and Regional Health Technology Assessment*, 2016.
- **Heath, A.**, Manlopoulou, I., Baio, G. “Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation." *Statistics in Medicine*, 35(23), 2016.

Papers in Preparation

- **Heath, A.**, and Baio, G. “Calculating the Expected Value of Sample Information using Efficient Nested Monte Carlo: A Tutorial" Submitted to *Value in Health*.
- **Heath, A.**, and Baio, G. and Hunter, R. “Development of a New Software Tool to Compute the Expected Value of Sample Information: An application to the HomeHealth intervention”. Presented at *HESG Winter Meeting 2018.*
- **Heath, A.**, Manolopoulou, I., and Baio, G. “Estimating the Expected Value of Sample Information Across Different Sample Sizes using Moment Matching and Non-Linear Regression"
- **Heath, A.**. Baio, G., Manolopoulou, I. and Welton, N. “Research Prioritisation with Value of Information Measures: A Decision Modelling Perspective"

Software

- Baio, G., and Beraldi, A. and **Heath, A.** “BCEA: Bayesian Cost Effectiveness Analysis”. Available from `CRAN` or `http://www.statistica.it/gianluca/BCEA`.
- **Heath, A.**, and Baio, G. “EVSI: A suite of functions for the calculation and presentation of the Expected Value of Sample Information”. Available from GitHub: `annaheath/EVSI`

Invited Presentations

- Health Economics @ Cambridge, Cambridge, UK — April 2018
- Health Economic Research Centre, Oxford, UK — Jan 2018
- PRIMENT Clinical Trial Unit, London, UK — Apr 2017
- Multi-Parameter Evidence Synthesis Group Seminar, Bristol, UK — Oct 2016
- 2nd Meeting for Value of Information Calculation Methods, Sheffield, UK — Oct 2015

Workshops

- ISPOR European Congress, Milan, Italy — Nov 2015
  With Baio, G., Welton, N., Strong M.

CONTENTS

1 INTRODUCTION TO VALUE OF INFORMATION 21
1.1 Notation and basic concepts 24
1.2 Probabilistic Sensitivity Analysis 25
1.3 Value of Information Measures 26
1.3.1 Expected Value of Perfect Information 26
1.3.2 Expected Value of Perfect Partial Information 28
1.3.3 Expected Value of Sample Information 29
1.4 Monte Carlo Calculation for Value of Information Measures 30
1.4.1 Calculating the EVPI by simulation 30
1.4.2 Calculating the EVPPI by MC simulation 31
1.4.3 Calculating the EVSI by MC simulation 31
2 CALCULATION METHODS FOR THE EXPECTED VALUE OF PARTIAL PERFECT INFORMATION 33
2.1 Finding analytical expressions for the inner integral 33
2.2 Approximations for Single Parameter EVPPI 35
2.2.1 The Strong and Oakley Method 35
2.2.2 The Sadatsafavi et al. Method 36
2.3 Non-parametric regression for EVPPI calculations 38
2.3.1 GAM regression within `mgcv` 39
2.4 Comparison of methods 40
2.4.1 Case Studies 40
2.4.2 Analysis 41
2.4.3 Comparison of current EVPPI estimation methods 43
2.4.4 Computational Time 47
2.5 Summary of EVPPI Calculation Methods 48
2.6 Fast Gaussian Process Regression 49
2.6.1 Hyperparameter estimation for GP regression 51
2.7 The Integrated Nested Laplace Approximation 52
2.7.1 The INLA algorithm 53
2.8 Spatial statistics and Stochastic Partial Differential Equations 53
2.8.1 Computing the EVPPI using SPDE-INLA 56
2.9 Testing the INLA-SPDE Gaussian Process Regression Method 57
2.9.1 Computational Time 58
2.10 Assessing the accuracy of the EVPPI estimators 59
2.10.1 Monotonicity with respect to the number of parameters of interest 59
2.10.2 The SPDE-INLA method as an EVPI approximation 61

2.10.3 Analytic Results 62
2.10.4 Standard Error 63
2.11 Conclusion 63
3 ESTIMATION METHODS FOR THE EXPECTED VALUE OF SAMPLE INFORMATION 65
3.1 Current Methods for EVSI estimation 65
3.2 Preposterior analysis 67
3.2.1 The distribution of the preposterior mean 68
3.2.2 Examples of the distribution of the preposterior mean 69
3.3 Estimating the distribution of the preposterior mean 72
3.3.1 Expectation and Variance for the preposterior mean 72
3.3.2 Moment Matching 73
3.4 Approximating the preposterior mean distribution by simulation 76
3.4.1 Estimating the variance of the preposterior mean 77
3.4.2 Calculating the EVSI for a specific set of treatment options 78
3.4.3 Nuisance Parameters 78
3.4.4 The Moment Matching algorithm for EVSI estimation 80
3.5 Examples for the EVSI estimation procedure 81
3.5.1 Discrete Samples with Beta-Binomial Conjugacy 81
3.5.2 Non-linear mean function with Exponential-Gamma conjugacy 83
3.5.3 Case Study: Ades et al. Decision Tree Model 84
3.6 Calculating the Variance of the preposterior mean 86
3.6.1 A new chemotherapy drug 89
3.6.2 Determining the optimal number of quadrature points 90
3.7 Summary of the basic moment matching method 91
3.8 Calculating the EVSI for different sample sizes 93
3.8.1 A Bayesian non-linear model 94
3.8.2 Calculating the EVSI for different sample sizes 94
3.9 Estimating this EVSI using non-linear regression 95
3.9.1 The Brennan and Kharroubi Example 96
3.9.2 A Health Economic Model for Chronic Pain 97
3.10 Estimating the EVSI across different sample sizes: results 99
3.10.1 The Brennan and Kharroubi example 99
3.10.2 Chronic Pain 99
3.11 Discussion of the EVSI calculation methods 103
4 PRACTICAL CONSIDERATIONS FOR GENERAL PURPOSE SOFTWARE 105
4.1 The `R` package `BCEA` 105
4.1.1 Using `BCEA` for the EVPI 106
4.2 Calculating the EVPPI using `BCEA` 107
4.2.1 Calculating the EVPPI for different willingness-to-pay values 107
4.2.2 Incremental Costs and Effects 107

4.2.3 The SPDE-INLA mesh 108
4.2.4 The INLA procedure 108
4.2.5 Principal Fitted Components 109
4.3 Linear Predictor and Residual Plotting for EVPPI estimation 110
4.4 Software to calculate the EVSI 112
4.4.1 Calculating the EVSI for different willingness-to-pay values 113
4.4.2 Sampling from the prior-predictive distribution 114
4.4.3 Calculating the EVSI in multi-decision problems 115
4.5 Visualisations for the EVSI 116
4.5.1 Displaying uncertainty in the EVSI estimate 116
4.5.2 Expected Net Benefit of Sampling 118
4.6 Conclusion 122
5 SUMMARY AND CONCLUDING REMARKS 123

LIST OF FIGURES

Figure 1 A visualisation of an EVPI calculation. 27

Figure 2 A Pictorial Description of the EVSI 29

Figure 3 An example of two plots showing the use of the visual tool for the Sadatsafavi *et al.* for calculating single parameter EVPPI. 38

Figure 4 The EVPPI values, plus or minus one standard error for the parameters in the Vaccine example using 5 different methods - Strong and Oakley single parameter estimation (purple - SO), Sadatsafavi *et al.* single parameter estimation (blue - SAD), Gaussian Process (green - GP) and GAM regression (pink - GAM) and two-step Monte Carlo simulation (red crosses - MC) 44

Figure 5 The EVPPI values, plus or minus one standard error for the parameters in the FEAST example using 5 different methods - Strong and Oakley single parameter estimation (purple - SO), Sadatsafavi *et al.* single parameter estimation (blue - SAD), Gaussian Process (green - GP) and GAM regression (pink - GAM) and two-step Monte Carlo simulation (red crosses - MC). Groups 1, 2 and 3 contain the following parameters respectively, (d_2^M, d_3^M, d_4^M, α, p^B, d^S, p^L), (d_2^M, d_3^M, d_4^M, α, p^B, d^S, q^M) and (d_2^M, d_3^M, d_4^M, d^S) 45

Figure 6 The computational time required for EVPPI estimation for the GP regression method, varying with the number of parameters for which the EVPPI is being calculated. 48

Figure 7 An example of the grid approximation used to approximate the Matérn GP in a spatial problem. The thick black line represents the border of Switzerland. The blue dots represent the positions where data points have been observed. These data points are used to estimate the value of the Matérn GP throughout the geographical space (i.e. the whole area covered by Switzerland, in this case). 55

Figure 8 The EVPPI estimate for the Gaussian Process regression method (GP) and the new method developed in this paper (SPDE) for increasing parameter subset size for the Vaccine (panel a) and the SAVI (panel b) case studies. 60

Figure 9 The distribution of the exact preposterior mean for different samples sizes using Exponential-Gamma conjugacy, with the prior for the net benefit marked in black. 71

Figure 10 The distribution of the incremental net benefit conditional on both the model parameters $\boldsymbol{\theta}$ (LHS) and conditional on the parameter of interest ϕ (RHS) for the simple two-parameter model. 79

Figure 11 The distribution of the EVSI estimator over 10 000 different simulations from the prior of θ for 4 different sample sizes for $\boldsymbol{X}$ for the Beta-Binomial conjugate model. The red line represents the analytical value of the EVSI. 82

Figure 12 The distribution of the EVSI estimator over $10\,000$ different simulations from the prior of θ for 4 different samples sizes for $\boldsymbol{X}$ using the Exponential-Gamma conjugate model. The red line represents the analytic value of the EVSI. 84

Figure 13 The "sampling distribution" of the EVSI conditional on the distribution over the different estimates for the variance of the preposterior mean for ϕ_1, ϕ_2 and ϕ_3 for the Ades *et al.* example [1]. The solid red line represents the EVSI calculated using Monte Carlo methods and 10^{10} simulations. The dashed blue line represents the EVSI estimate obtained using the Strong *et al.* method with 10^6 simulations. The dotted brown line represents the average EVSI estimate obtained using moment matching and the solid black horizontal line represents the 90% interval for this estimation method. The comparator methods (MC and Strong *et al.*) are taken from Strong *et al.* [112] for the first three graphics. 87

Figure 14 The estimate of the variance of the preposterior mean for increasing numbers of quadrature points. The red line gives the variance of the preposterior mean calculated using all the samples in the prior for θ. The dashed lines are the standard errors for the estimates of the variance of the preposterior mean. 88

Figure 15 The sampling distribution of the EVSI estimates calculated using variance estimates with different numbers of quadrature points and a total of $500\,000$ posterior simulations. The number of quadrature points for the variance estimate are given in the title. The vertical line represents the MC estimate for the EVSI. 91

Figure 16 The sampling distribution of the EVSI estimates calculated using variance estimates with different numbers of quadrature points and a total of $5\,000$ posterior simulations. The number of quadrature points for the variance estimate are given in the title. The vertical line represents the MC estimate for the EVSI. 92

Figure 17 The EVSI estimate with error bounds for data collection exercise 1 (top) and exercise 2 (bottom) in the BK example. These are compared with the 2-level MC estimates (red dots) and the Menzies estimates (blue crosses). 100

Figure 18 The EVSI estimate with error bounds for data collection exercise 3 (top) and exercise 4 (bottom) in the BK example. These are compared with the 2-level MC estimates (red dots) and the Menzies estimates (blue crosses). 101

Figure 19 The EVSI estimate with error bounds for data collection exercise 5 in the BK example. These are compared with the 2-level MC estimates (red dots) and the Menzies estimates (blue crosses). 102

Figure 20 The economic value of the two alternative questionnaire collection strategies for different sample sizes in the chronic pain example. The grey curves indicate the potential value of the study without incentives and the black curves the value of the study with incentives estimated using the moment matching method. The black crosses represent the no-incentive MC simulation estimates and the grey crosses represent the same estimates for the incentive study. 103

Figure 21 Expected value of perfect information for the Vaccine example §2.4.1.1. 106

Figure 22 Two grid approximations for the same data set. The LHS shows the triangulation when the variables are left on their original scale, with the projected data points in blue. Notice that there are a large number of triangles in this case, but a relatively small number that surround the data points. In contrast to this, on the right, where the data points are scaled we note that a much larger number of mesh points cover the data, allowing for a more accurate Matérn field approximation for a fixed computational time. 109

Figure 23 Residual plots for the incremental costs and effects for treatment 3 and 4 in the FEAST example with no interactions in the linear predictor. Treatment 1 is taken as the reference treatment. 111

Figure 24 Residual plots for the incremental costs and effects for treatment 3 and 4 in the FEAST example with 3rd order interactions in the linear predictor. Treatment 1 is taken as the reference treatment. 112

Figure 25 An example of the graph produced when an `evppi` object is plotted in `R`. 113

Figure 26 The EVSI plotted across different willingness-to-pay thresholds, (a) shows the EVSI for one sample size and (b) across a large number of sample sizes. 117

Figure 27 The EVSI plotted across different values of N for a specified willingness-to-pay value. The graphic also shows uncertainty bands for the EVSI. 118

Figure 28 The probability of a cost-effective trial for different lengths for the time horizon and levels of incidence population. The probability represents the uncertainty in the study costs and EVSI estimates. 120

Figure 29 The ENBS across different sample sizes. The alternative curves represent the uncertainty in the value of the ENBS arising from uncertainty in knowledge about the study costs and the EVSI estimation procedure. 121

LIST OF TABLES

Table 1 A PSA "dataset" containing the simulated parameter values for all three parameters in this model, along with the Net Benefit values for two treatments 26

Table 2 A PSA "dataset" containing the simulated parameter values for all three parameters in this model, along with the Net Benefit values for two treatments, extended to demonstrate how to estimate the EVPI by MC simulation. 31

Table 3 The computational time required (in seconds) to calculate an EVPPI using both the GP regression method and SPDE-INLA method for increasing numbers of parameters for both case studies 59

Table 4 The EVPI values calculated using the PSA samples directly and our SPDE-INLA method 62

Table 5 Comparison of the EVPPI estimation methods, standard GP, GAM regression and the SPDE-INLA method with "true" EVPPI values based on 10^7 Monte Carlo simulations. 62

Table 6 The EVSI estimate for different numbers of posterior samples using the moment matching method. 89

Table 7 The single parameter EVPPI values for all the parameters in the Chemotherapy model. 90

Table 8 The Parameters for the Brennan and Kharroubi example. The mean and standard deviations for the distributions of the parameters is also given. 96

Table 9 The assumed standard deviation for the data collection exercise targeting uncertainty in the corresponding parameter. 97

ACRONYMS

AE Adverse Effects

BCEA Bayesian Cost Effectiveness Analysis

ENBS Expected Net Benefit of Sampling

EVPI Expected Value of Perfect Information

EVPPI Expected Value of Perfect Partial Information

EVSI Expected Value of Sample Information

FEAST Trial acronym for the Fluid Expansion as Supportive Therapy trial

GAM Generalised Additive Model

GMRF Gaussian Markov Random Field

GP Gaussian Process

HTA Health Technology Assessment

INB Incremental Net Benefit

INLA Integrated Nested Laplace Approximations

LHS Left Hand Side

MC Monte Carlo

MCMC Markov Chain Monte Carlo

MINT Trial acronym for the Managing Injuries of the Neck Trial trial

NB Net Benefit

NICE National Institute of Health and Care Excellence

NS Neurological Sequelae

PFC Principal Fitted Components

PSA Probabilistic Sensitivity Analysis

QALY Quality Adjusted Life Years

RHS Right Hand Side

SAVI Sheffield Accelerated Value of Information

SPDE Stochastic Partial Differential Equation

VoI Value of Information

1

INTRODUCTION TO VALUE OF INFORMATION

Broadly speaking, the objective of health economic evaluations is to maximise health gains across a population given finite monetary resources and limited budget. These evaluations are, therefore, the basis of assessments by bodies such as the National Institute for Health and Care Excellence (NICE) in the UK that provide guidance on decision-making for publicly funded health-care systems. They encompass a suite of approaches for combining costs and clinical consequences of an intervention, in order to compare the intervention with alternative options which may already be available. Typically, an assessment of the impact of uncertainty on the decision making process is also included. Therefore, health-economics is, arguably, a branch of applied statistics [13, 121], which is often performed using the Bayesian approach [4, 90, 91, 106].

The process of health-economic evaluations involves the identification of suitable measures of clinical benefits (generically termed as "effectiveness") and costs associated with an intervention. The costs usually include the cost of acquisition and implementation of the health intervention (e.g. a drug), along with societal costs such as those related to number of days off work or social care. As for the clinical benefits, they can be a "hard" measurement (e.g. number of cases averted), but are, most often, considered in terms of *Quality Adjusted Life Years* (QALYs) [75], combining the quantity and the quality of life provided by a given intervention. Individual level variability in the costs and effectiveness is normally expressed in terms of a joint probability distribution, indexed by a set of population level parameters.

Once the modelling step is complete and all relevant costs and benefits are included in the model, the decision making process is performed under a strict decision theoretic framework [96]. For each intervention, the health economic outcomes are combined into a *utility function*, which quantifies the overall "value" of the intervention. The alternative associated with the highest expected utility is then deemed as the most cost-effective, given current evidence — notice that in a Bayesian context, this expectation is taken over the distributions of both the individual level outcomes and population level parameters. From the decision-theoretic point of view, the identification of the overall expected utility is all that is needed to reach the best decision given the current state of knowledge available to the decision-maker [24, 25].

However, the implementation of a health care intervention is usually associated with risks such as the irreversibility of investments [23]. Moreover, health economic models often involve a relatively large number of parameters, usually estimated with limited information. For these reasons, a thorough investigation of the impact of uncertainty on the decision making

process of parametric uncertainty is recommended by several Health Technology Assessment (HTA) bodies, [17, 28, 38, 45]. This process is known in the health economics literature as *Probabilistic Sensitivity Analysis* (PSA).

Typically, PSA is carried out by defining probability distributions for the population parameters which represent the current level of uncertainty in these parameters. These distributions then define a distribution for the costs and effectiveness measures and can therefore be used to determine a distribution over the different possible decisions based on the current uncertainty level in the model parameters. In most settings, this distribution is summarised by estimating the probability that each treatment option is cost effective. However, this analysis can dramatically under/overestimate the sensitivity of the decision to uncertainty. For example, if all treatment options have similar costs and effectiveness, there will be low probability of cost effectiveness as the comparators are almost equally likely to be cost effective. However, agonising over the decision between the alternative treatments is a waste of resources as they all give approximately the same benefits and costs.

Therefore, the analysis of the *Value of Information* (VoI) [64] has become an increasingly popular method for conducting PSA in health economic evaluations [1, 9, 13, 23, 27, 46, 47, 50]. The basic idea of this analysis to compare the decision based on current evidence to the one that would be made had “more” information been available. In some settings, we consider that it is possible to resolve *all* the uncertainty in the model parameters, while others consider different strategies for collecting this additional information.

The main advantage of the analysis of the VoI over simply calculating the probability of cost effectiveness is that it directly takes into account the opportunity cost of making the wrong decision as well as the likelihood of modifying the decision. If this cost is low, the decision can be taken immediately, even for a low probability of cost-effectiveness as deferring the decision will have little value. Additionally, the value of the information can be directly compared with the cost of gathering additional information to reduce uncertainty which allows us to determine which studies should be funded. For this reason, it has been advocated that VoI measures should be presented when representing decision uncertainty [4, 16, 46, 83].

Despite these useful features, the uptake of VoI analysis in health economic evaluation has been slow. VoI analysis has been hindered by several different factors, both theoretical and practical. Theoretically, decision makers often find these measures more challenging to comprehend than the probability of cost-effectiveness [117]. This is partially to do with the fact that VoI measures are unbounded and therefore is it more challenging to determine which constitutes a “high” level of decision uncertainty, particularly when considering the value of resolving *all* model uncertainty.

More importantly, however, there are practical barriers to using VoI measures as they are very computationally costly unless restrictive assumptions are made [123]. It is relatively easy to calculate numerically (but not analytically) the expected value of learning *all* the model parameters perfectly. This is known as the (overall) *Expected Value of Perfect Information* (EVPI). However, this quantity often has little practical use (except in models where this value is very low), as it will be rarely possible to learn the *exact* value of all the underlying model

parameters. Thus, the decision-maker is usually interested in the expected value of learning about a subset of the model parameters or even more importantly the expected value of information from a specific study.

In practice, subset analysis indicates which parameters are driving decision uncertainty and is based on a VoI measure known as the *Expected Value of Perfect Partial Information* (EVPPI). This measures the value of learning the exact value of a specific subset of the model parameters. In general, the EVPPI can be calculated for different subsets of the model parameters to determine which parameters have the greatest economic value. These should then be targeted in a future study in order to efficiently reduce decision uncertainty. It is also possible to calculate the value of a proposed study with a specific design using a VoI measure known as the *Expected Value of Sample Information* (EVSI). Traditionally, both these measures have been highly computationally costly, with the EVSI more challenging to compute than the EVPPI.

Recently, significant effort has been invested in developing more efficient methods for computing the EVPPI [12, 46, 77, 86, 104, 110, 111] and the EVSI [1, 10, 11, 68, 70, 120]. Despite these advances, there are still limitations to all of these methods which continue to hinder the implementation of VoI measures in practical health economic evaluations. Therefore, this thesis is concerned with addressing some of these issues and then proposes new methodology to calculate both the EVPPI and the EVSI.

In the remainder of this chapter, a formal introduction to the health economic modelling framework and VoI measures is presented. This will include a discussion of the key concepts and notation that will be used throughout to present the VoI methods and their calculation methods. It will also discuss some key properties of these measures.

Chapter 2 follows this introduction by discussing calculation methods for the EVPPI. Firstly, a full literature review is undertaken to determine the current methods available to calculate the EVPPI. Two case studies are introduced to compare these methods and highlight the remaining difficulties associated with calculating the EVPPI. Following from this, §2.6 presents the methodology that we have developed to calculate the EVPPI efficiently. This methodology is based on a fast approximation to Gaussian Process regression and thus we begin this discussion by introducing Gaussian Process regression. We then present several different results that will allow us to build up a fast Gaussian Process approximation method. The methodology is then tested using two case studies to determine the accuracy and computational speed up.

Chapter 3 then presents calculation methods for the EVSI. We begin with a short discussion of the most recent calculation methods for the EVSI to motivate the development of our calculation method. Our method is then presented in simple terms before being extended to calculate the EVSI across different sample sizes for the proposed trial using Bayesian non-linear regression. Several technicalities relating to the method are also discussed and the method is tested using several different examples ranging in complexity before being applied to a real-life health economic model developed for assessing painkillers to treat chronic pain [113].

Chapter 4 then discusses the technical considerations required to integrate both these methods into general purpose software. We also discuss the graphics that have been developed to present both the EVPPI and EVSI. These graphics aid understanding and presentation of these VoI measures to ensure that they can be used in practical health economic evaluations. The software also allows practitioners to determine when they have issues with the estimation procedures. Therefore, the work presented in this thesis has addressed not only the major computational concerns regarding the use of VoI measures but has also addressed some of the theoretical concerns surrounding VoI measures by aiding practitioners to understand VoI measures more clearly through the use of standardised graphics.

1.1 NOTATION AND BASIC CONCEPTS

Health economic decision making is concerned with determining the *optimal* intervention in terms of costs and health effects among a number alternative options. In general, we compare $T+1$ alternative treatments, where the treatment with the index $t=0$ is the current standard of care that is being compared with T alternative, "innovative" treatments. Theoretically, it is possible to consider the case where a large number of alternative interventions is available, although in practice it is uncommon that $T>5$.

As previously stated, health economic modelling involves estimating the effectiveness and costs for each treatment under consideration, denoted by the pair (e_t, c_t). To determine the optimal treatment, these two health economic outcomes must be combined into a *utility function* to give a value for each of the treatments. In health economic evaluations, the treatments are typically defined in terms of the monetary net benefit [107]

$$nb_t = ke_t - c_t.$$

Here, (e_t, c_t) are typically subject to individual variability expressed by a joint probability distribution $p(e_t, c_t \mid \boldsymbol{\theta})$ conditional on the model parameters $\boldsymbol{\theta}$. The parameter k is known as the *willingness-to-pay* parameter. This value is typically set externally by the HTA body and represents the amount of money the body is willing to spend to extend someone's life by 1 year in perfect health. In England and Wales, NICE states that this threshold is between $£20\,000 - £30\,000$ [81] although recent analysis has estimated it closer to $13\,000$ [26].

In a full Bayesian setting, a complete ranking of the possible treatment options is obtained by computing the overall expectation of this utility function over both individual variability and parameter uncertainty

$$\mathcal{NB}_t = k\mathsf{E}[e_t] - \mathsf{E}[c_t],$$

i.e. the expectation here is taken with respect to the joint distribution $p(e, c, \boldsymbol{\theta}) = p(e, c \mid \boldsymbol{\theta})$ $p(\boldsymbol{\theta})$. The option t associated with the maximum overall expected utility $\mathcal{NB}^* = \max_t \mathcal{NB}_t$ is deemed to be the most "cost-effective", given current evidence. At this point, it is important to note that the ranking of the treatments is based on the willingness-to-pay values which is typically defined as an interval and not as a fixed number. Therefore, measures based on the

utility function should be calculated for different values of k. However, as the utility function is linear in k it is easy to calculate the expected monetary net benefit for many different values and therefore perform some level of sensitivity analysis to this willingness-to-pay threshold.

1.2 PROBABILISTIC SENSITIVITY ANALYSIS

In general, health economic evaluations are not simply concerned with the *optimal* treatment under current information but also with the impact of uncertainty on the decision making procedure. In this setting, the uncertainty in the model parameters is defined using the joint distribution $p(\boldsymbol{\theta})$, either using a fully Bayesian model or frequentist bootstrapping. This process is known as Probabilistic Sensitivity Analysis (PSA) and is concerned with understanding the impact of parametric uncertainty on the decision. Therefore, we are not interested in understanding the impact of individual level variation on the decision and so we consider the monetary net benefit as a function of $\boldsymbol{\theta}$ only by averaging out individual level uncertainty:

$$\mathrm{NB}_t(\boldsymbol{\theta}) = k\mathrm{E}[e_t \mid \boldsymbol{\theta}] - \mathrm{E}[c_t \mid \boldsymbol{\theta}]. \tag{1}$$

In line with [4], we term this the "known-distribution" net benefit. In the above expression, therefore, we are taking expectation with respect to the conditional distribution $p(e_t, c_t \mid \boldsymbol{\theta})$. Thus, while decision making is concerned with $\mathcal{NB}_t$, a deterministic quantity, PSA is concerned with the random variable $\mathrm{NB}_t(\boldsymbol{\theta})$ where $\mathrm{E}\left[\mathrm{NB}_t(\boldsymbol{\theta})\right] = \mathcal{NB}_t$.

Typically PSA is carried out using a simulation based approach [2, 4, 7]. In a Bayesian setting, models are normally estimated using a simulation approach, e.g. MCMC and therefore the known-distribution net benefit values are created as a by-product of the health-economic analysis. Thus, in a Bayesian setting, the simulated values for $\boldsymbol{\theta}$ come from posterior draws and the known-distribution net benefits are posterior draws from the distributions of the net benefits which are either calculated deterministically or stochastically from the underlying parameters. In a frequentist setting, PSA is performed in addition to the main base case analysis by using bootstrapping to obtain parameter samples, with the known-distribution net benefit values then calculated as a function of these parameter draws.

Irrespective of the setting in which these parameter samples are taken, we denote $\boldsymbol{\theta}_s$, $s = 1 \ldots S$, as a draw from the "distribution" of $\boldsymbol{\theta}$. We then denote the known-distribution net benefit for this parameter vector as $\mathrm{NB}_t(\boldsymbol{\theta}_s)$. Note, therefore, that throughout this thesis a "dataset" and "data" with the s subscript do not typically represent data in the traditional sense. They are simply simulated values from distributions representing the uncertainty in the parameter estimates. For further clarity, Table 1 demonstrates this for the simple example, where the underlying model is characterised by three parameters, $\boldsymbol{\theta} = (\pi_1, \pi_2, \rho)$.

From this "dataset", it is straightforward to calculate the probability of cost-effectiveness for each of the treatments under consideration; simply calculate the proportion of simulations for which the treatment is considered optimal. Therefore, it is trivial to use the probability of cost-effectiveness as a summary measure for PSA [28, 49].

Simulation	π_1	π_2	ρ	NB_1	NB_2	$\max_t$
1	2875	55001	8.7	3249597	4313158	NB_2
2	3617	76051	5.7	6269563	5594843	NB_1
3	4312	32016	9.1	4706446	6895321	NB_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
S	2161	45791	7.6	2797630	3253655	NB_2

Table 1: A PSA "dataset" containing the simulated parameter values for all three parameters in this model, along with the Net Benefit values for two treatments

1.3 VALUE OF INFORMATION MEASURES

To define the VoI measures, recall that the optimal decision under current information is the treatment that maximises the expected known-distribution net benefit,

$$\max_t \mathsf{E}_{\boldsymbol{\theta}}\left[\text{NB}_t(\boldsymbol{\theta})\right] = \max_t \mathcal{NB}_t = \mathcal{NB}^*.$$

Broadly speaking, VoI measures assume that it would be possible to gain extra information that changes the known-distribution net benefit. In turn, this information *could* change the optimal decision and the newly optimal treatment would give an alternative value for the optimal decision. The value of the information is then the difference between the value of the current optimal decision and the value of the decision made with this additional information. Alternative VoI measures then arise by considering different potential methods for collecting the additional information.

1.3.1 *Expected Value of Perfect Information*

First of all, consider that it is possible to learn the *exact* value of every model parameter. This is known as learning "perfect information" about $\boldsymbol{\theta}$ and to find its value, consider the value of learning that $\boldsymbol{\theta}$ is exactly equal to $\boldsymbol{\theta}'$. In this case, the value of the optimal decision would be

$$\max_t \text{NB}_t(\boldsymbol{\theta}'),$$

where it is not necessary to take expectations as there is no remaining uncertainty in the model (individual level variability is already marginalised out).

The above quantity is the value of learning that $\boldsymbol{\theta} = \boldsymbol{\theta}'$; however, as the value of the underlying model parameters is not known, we need to summarise the value of the information across all possible values of $\boldsymbol{\theta}'$. This is achieved by taking the expectation over all values of $\boldsymbol{\theta}'$ to calculate the Expected Value of Perfect Information (EVPI);

$$\text{EVPI} = \mathsf{E}_{\boldsymbol{\theta}}\left[\max_t \text{NB}_t(\boldsymbol{\theta})\right] - \max_t \mathsf{E}_{\boldsymbol{\theta}}\left[\text{NB}_t(\boldsymbol{\theta})\right]. \tag{2}$$

To comprehend the EVPI more clearly, consider the visualisation in Figure 1. Two treatments are under consideration and the known-distribution net benefit for each treatment is defined as a deterministic function of a single model parameter θ. If the true value of θ is less than 0.4 then treatment 1 is optimal whereas if θ is greater than 0.4 then treatment 2 is optimal. The expected values of each treatment under current information are represented by the purple and blue dots on the graph for treatments 1 and 2 respectively. Therefore, the current optimal treatment is treatment 2 and its value is represented by the blue dot and gives the second term in equation (2).

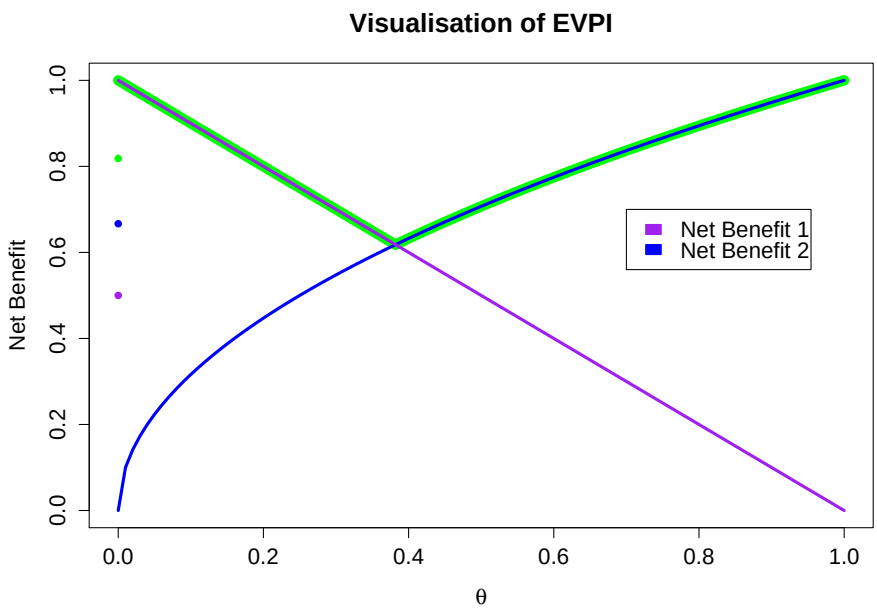


Figure 1: A visualisation of an EVPI calculation.

However, the EVPI is calculated by finding the value of the optimal treatment at each value of θ, represented by the green line, and then taking the average, represented by the green dot. This quantity is the first term in equation (2). Therefore, in this example, the EVPI is then defined as the difference between the green and blue dots. In general, this quantity is non-negative as the current *optimal* decision only changes when the net benefit of another treatment dominates the globally *optimal* treatment, as for $\theta < 0.4$ in Figure 1.

The EVPI gives an upper bound for any study aimed at reducing uncertainty in any of the model parameters. If this value is very small, then it is possible to conclude that there is little value in investigating any of the model parameters and therefore the decision on the optimal treatment can be taken using the current information without considering collecting additional data. If, however, the EVPI shows that there is value in a future study, it is useful to determine which parameters are driving decision uncertainty and therefore where future research should be targeted.

1.3.2 *Expected Value of Perfect Partial Information*

We begin by assessing the value of obtaining perfect information about a specific subset of parameters. Consider that the parameter vector can be split into two components $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\psi})$, where $\boldsymbol{\phi}$ is the sub-vector of parameters of interest (i.e. those that could be investigated further) and $\boldsymbol{\psi}$ are the remaining "nuisance" parameters, sometimes known as *focal* and *non-focal* parameters respectively. The VoI measure that gives the value of obtaining perfect information about $\boldsymbol{\phi}$ is known as the Expected Value of Perfect Partial Information (EVPPI) and is computed in a similar manner to the EVPI. Firstly, assume that we have perfect information about the parameters of interest $\boldsymbol{\phi} = \boldsymbol{\phi}'$. Under this condition, the value of the optimal treatment is then equal to

$$\max_t \mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}'} \left[\mathsf{NB}_t(\boldsymbol{\phi}', \boldsymbol{\psi})\right], \tag{3}$$

where the expectation is taken over the remaining uncertainty in the nuisance parameters.

The EVPPI is then calculated by taking the expectation of equation (3) across the support of $\boldsymbol{\phi}$ and subtracting the value of the current optimal decision:

$$\mathsf{EVPPI} = \mathsf{E}_{\boldsymbol{\phi}} \left[\max_t \mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}} \left[\mathsf{NB}_t(\boldsymbol{\phi}, \boldsymbol{\psi})\right]\right] - \max_t \mathsf{E}_{\boldsymbol{\phi},\boldsymbol{\psi}} \left[\mathsf{NB}_t(\boldsymbol{\phi}, \boldsymbol{\psi})\right]. \tag{4}$$

As with the EVPI, the EVPPI is bounded below by 0. It is also bounded above by the EVPI and more generally, if there are two parameter subsets $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$ such that $\boldsymbol{\xi} \subset \boldsymbol{\phi}$, the EVPPI of $\boldsymbol{\phi}$ is greater than or equal to the EVPPI of $\boldsymbol{\xi}$.

To demonstrate this, assume that $\boldsymbol{\phi} = (\boldsymbol{\xi}, \boldsymbol{\xi}^c)$, meaning that $\boldsymbol{\xi}^c$ is the set of "additional" nuisance parameters arising within $\boldsymbol{\phi}$:

$$\begin{aligned}
\mathsf{EVPPI}(\boldsymbol{\phi}) &= \mathsf{E}_{\boldsymbol{\phi}} \left[\max_t \mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right]\right] - \max_t \mathsf{E}_{\boldsymbol{\theta}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right] \\
&= \mathsf{E}_{\boldsymbol{\xi}} \left[\mathsf{E}_{\boldsymbol{\xi}^c|\boldsymbol{\xi}} \left[\max_t \mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right]\right]\right] - \max_t \mathsf{E}_{\boldsymbol{\theta}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right] \\
&\geq \mathsf{E}_{\boldsymbol{\xi}} \left[\max_t \mathsf{E}_{\boldsymbol{\xi}^c|\boldsymbol{\xi}} \left[\mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right]\right]\right] - \max_t \mathsf{E}_{\boldsymbol{\theta}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right] \\
&= \mathsf{E}_{\boldsymbol{\xi}} \left[\max_t \mathsf{E}_{\boldsymbol{\xi}^c|\boldsymbol{\xi}} \left[\mathsf{E}_{\boldsymbol{\psi}|(\boldsymbol{\xi},\boldsymbol{\xi}^c)} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right]\right]\right] - \max_t \mathsf{E}_{\boldsymbol{\theta}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right] \\
&= \mathsf{E}_{\boldsymbol{\xi}} \left[\max_t \mathsf{E}_{(\boldsymbol{\psi},\boldsymbol{\xi}^c)|\boldsymbol{\xi}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right]\right] - \max_t \mathsf{E}_{\boldsymbol{\theta}} \left[\mathsf{NB}_t(\boldsymbol{\theta})\right] = \mathsf{EVPPI}(\boldsymbol{\xi})
\end{aligned}$$

by Jensen's inequality as the function $\max(\cdot)$ is convex. While this property is logical, it is always more valuable to learn about a larger number of parameters, it is also an important tool for assessing EVPPI estimation methods and so must be demonstrated formally.

In a similar manner to the EVPI, the EVPPI is most useful when it demonstrates that there is little value in learning about a subset of parameters. This is because it is rarely possible to obtain perfect information about a model parameter. Therefore, if the EVPPI for a set of

parameters is high, then it is of interest to determine how the uncertainty in these parameters should be reduced.

1.3.3 *Expected Value of Sample Information*

The value of a specific study, aimed at reducing uncertainty in a subset of the model parameters, is known as the Expected Value of Sample Information (EVSI). It is defined in a similar way to both the EVPI and the EVPPI, but rather than learning the exact value of the parameters we consider that the study would give rise to data X. These data update the information about the model parameters in a Bayesian manner and the value of the optimal decision under this additional information is

$$\max_t \mathrm{E}_{\theta|X}\left[\mathrm{NB}_t(\theta)\right], \tag{5}$$

where $\mathrm{E}_{\theta|X}\left[\mathrm{NB}_t(\theta)\right]$ is the posterior mean of the known-distribution net benefit.

Again, as it is not known what data will arise from the future study, we calculate the EVSI as the expectation of (5) over all the possible data sets from the future trial;

$$\mathrm{EVSI} = \mathrm{E}_X\left[\max_t \mathrm{E}_{(\theta|X)}\left[\mathrm{NB}_t(\theta)\right]\right] - \max_t \mathrm{E}\left[\mathrm{NB}_t(\theta)\right], \tag{6}$$

where the distribution of the future data $X \sim p(X)$ is the prior predictive distribution of the data. In general, the distribution for the data is defined conditional on the model parameters through $p(X \mid \theta)$ and then the parameter uncertainty is integrated out;

$$p(X) = \int_{\Theta} p(X \mid \theta)p(\theta)d\theta,$$

where $p(\theta)$ is the prior/PSA distribution for the model parameters θ.

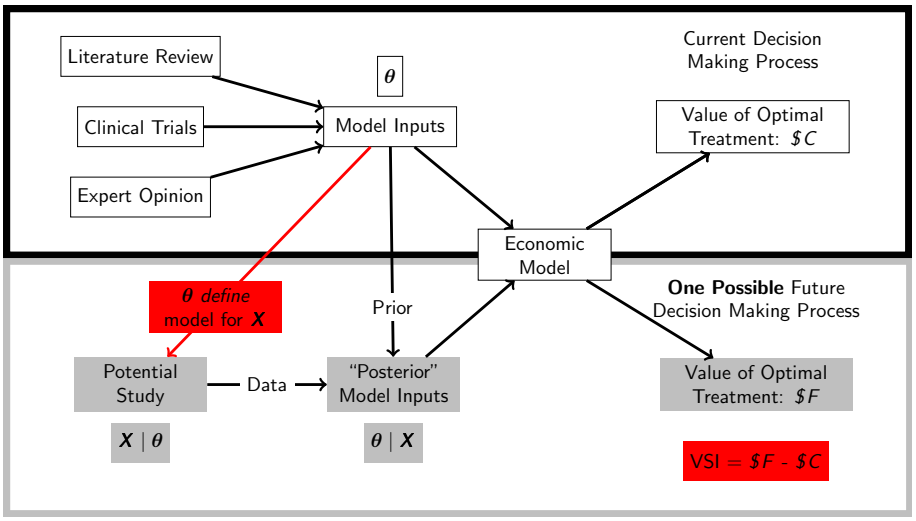


Figure 2: A Pictorial Description of the EVSI

To aid comprehension, Figure 2 gives a pictorial representation of the EVSI. The top half of the picture represents the standard health economic decision making process, where the

distribution of the model parameters $\boldsymbol{\theta}$ are dependent on synthesizing different information sources. The model parameters are then fed through the economic model to determine the known-distribution net benefit and the optimal treatment is found by taking the expectation of the known-distribution net benefit. To calculate the EVSI, a potential study is designed by determining a relationship between the model parameters and the future data. The data from this study is used to update the model parameters and then a new distribution for the known-distribution net benefit is found using the economic model. Finally, the optimal treatment conditional on this new dataset is found and the value of the sample information is the difference between these two values. The EVSI is then calculated as the average over all the potential datasets.

Finally, the EVSI can be compared directly with the cost of the study to determine whether there is net value in undertaking the study. If the EVSI is greater than the cost of the study, then it should go ahead. If, on the other hand, the study cost exceeds the EVSI, then the current information is sufficient to make a decision. Therefore, we define the Expected Net Benefit of Sampling (ENBS) as the EVSI minus the study costs. The ENBS can then be used to make decisions about whether the study should be undertaken or not. The ENBS can also be used to design trials as we can search for the trial with the largest ENBS, further discussion of the ENBS is given in §4.5.2.

1.4 MONTE CARLO CALCULATION FOR VALUE OF INFORMATION MEASURES

As the three key VoI measures have been defined, we consider methods for estimating these quantities. Theoretically, these quantities are based on computing expectation and so it could be possible to calculate them using integration. However, the integral of a maximum is very challenging to compute even in simple settings implying that VoI measures are rarely available analytically. However, it is possible to calculate all the measures using Monte Carlo (MC) simulation, although, as will be demonstrated, this is very computationally intensive for the EVPPI and the EVSI.

1.4.1 *Calculating the EVPI by simulation*

Once the PSA dataset is available, it is trivial to calculate the EVPI as the estimate relies directly on the known-distribution net benefit values. Consider the definition of the EVPI in (2), we can see that the MC estimator of the EVPI is,

$$\widehat{\text{EVPI}} = \frac{1}{S}\sum_{s=1}^{S} \max_t \text{NB}_t(\boldsymbol{\theta}_s) - \max_t \frac{1}{S}\sum_{s=1}^{S} \text{NB}_t(\boldsymbol{\theta}_s).$$

Revisiting the PSA dataset in Table 2, we can see that the EVPI is calculated by taking the row-wise maximum for each simulation and then averaging over all the rows, giving a value of $4\,551\,204$. The EVPI is then calculated as the difference between this value and the value of the optimal treatment ($3\,253\,655$).

Simulation	π_1	π_2	ρ	NB_1	NB_2	$\max_t [\text{NB}_t]$	$\max_t$
1	2875	55001	8.7	3249597	**4313158**	4313158	NB_2
2	3617	76051	5.7	**6269563**	5594843	6269563	NB_1
3	4312	32016	9.1	4706446	**6895321**	6895321	NB_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
S	3616	4762	6.7	**5629348**	2301928	5629348	NB_1
Mean	2161	45791	7.6	2797630	3253655	4551204	NB_2

Table 2: A PSA "dataset" containing the simulated parameter values for all three parameters in this model, along with the Net Benefit values for two treatments, extended to demonstrate how to estimate the EVPI by MC simulation.

1.4.2 *Calculating the EVPPI by MC simulation*

Calculating the EVPPI by MC simulation is more challenging and cannot be based directly on the PSA samples. This is due to the inner expectation in the first term in equation (4). Strictly, the MC estimation method for the EVPPI was first formalised in [12] and it uses nested simulation to estimate the conditional expectation $\text{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}[\text{NB}_t(\boldsymbol{\theta})]$. To estimate the EVPPI, R values must be sampled from the conditional distribution of $\boldsymbol{\psi} \mid \boldsymbol{\phi}_s$ for every PSA sample $\boldsymbol{\phi}_s$, $s = 1 \ldots S$. These R values are then used to calculate the known-distribution net benefit with $\boldsymbol{\phi} = \boldsymbol{\phi}_s$ and then to find the treatment with the maximum expected net benefit:

$$\widehat{\text{EVPPI}} = \frac{1}{S}\sum_{s=1}^{S}\left[\max_t \frac{1}{R}\sum_{r=1}^{R}[\text{NB}_t(\boldsymbol{\phi}_s, \boldsymbol{\psi}_r)]\right] - \max_t \frac{1}{S}\sum_{s=1}^{S}\text{NB}_t(\boldsymbol{\theta}_s). \tag{7}$$

The main difficulty of using this estimator is the computational time required for this nested simulation strategy. This method requires $S \times R$ computations, where S and R can typically be in the order of $10\,000$ [12]. In addition to the sheer volume of simulations, some economic models can be expensive to run making this MC estimator impractical and occasionally impossible to implement within realistic time frames. Nonetheless, as this estimator derives directly from the definition of the EVPPI, we consider that this MC estimator with a long simulation run is the "gold-standard" estimator for the EVPPI.

Note that, for low dimensional $\boldsymbol{\phi}$, it is possible to reduce the computation time for this estimator slightly by using quadrature to estimate the outer expectation [35]. This allows us to cover the parameter space intelligently and therefore significantly reduce the number of "inner" simulations, i.e. simulations from $\boldsymbol{\psi} \mid \boldsymbol{\phi}_s$, requiring around 30-50 rather than S simulations from the conditional distribution. However, for higher dimensional vectors $\boldsymbol{\phi}$ this quadrature method becomes more expensive than MC sampling.

1.4.3 *Calculating the EVSI by MC simulation*

Finally, calculating the EVSI by MC simulation follows a similar strategy to the EVPPI, although an extra simulation step is needed to simulate the potential data sets. As the distribution for

the potential data is defined using $p(\boldsymbol{X} \mid \boldsymbol{\theta})$, the PSA values for the model parameters should be used to estimate a potential data set for each PSA simulation, i.e. a sample $\boldsymbol{X}_s$ must be simulated from $p(\boldsymbol{X} \mid \boldsymbol{\theta}_s)$ for $s = 1, \ldots, S$. Conditional on each sample, R simulated parameter values are then sampled from the posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{X}_s)$. These R parameter values are used to calculate the known-distribution net benefit and therefore the value of the optimal treatment under sample information:

$$\widehat{\text{EVSI}} = \frac{1}{S}\sum_{s=1}^{S}\left[\max_t \frac{1}{R}\sum_{r=1}^{R}\left[\text{NB}_t\left(\boldsymbol{\theta}_r \mid \boldsymbol{X}_s\right)\right]\right] - \max_t \frac{1}{S}\text{NB}_t\left(\boldsymbol{\theta}_s\right). \tag{8}$$

Clearly this estimator has a similar computational cost to the EVPPI estimator as $R \times S$ model runs are required. However, as the posterior distribution for the model parameters will normally need to be estimated using MCMC methods rather than MC simulations, the inner simulations normally have a higher computational cost than those for the EVPPI. This is because MCMC methods require burn-in and potentially thinning for all but the simplest models. In addition to this, as the EVSI is concerned with the value of a specific study, the EVSI may need to be calculated for multiple alternative study designs. This introduces additional complexity to an already very computationally intensive procedure. Therefore, there is a need for computationally efficient estimators for both the EVPPI and the EVSI before VoI measures can be used in health economic evaluations.

The following two chapters present the computationally efficient estimation methods for both the EVPPI and EVSI. Both chapters begin with a presentation of the current available calculation methods before moving on to our two novel calculation methods that have been developed. These new methods are tested in a variety of health economic models and compared in terms of accuracy and computational time with the most recent innovations in this field.

2

CALCULATION METHODS FOR THE EXPECTED VALUE OF PARTIAL PERFECT INFORMATION

This chapter begins with an extensive literature review detailing all the current methods available for calculating the EVPPI. All the methods are simulation based to some extent due to the difficulty of finding analytically tractable solutions for the expectation of a maximum. This literature review highlights the remaining issues with these EVPPI calculation methods and motivates the latter half of the chapter which presents a novel calculation method for the EVPPI. This new calculation method is based on non-parametric regression, specifically Gaussian Process regression [88, 98] and uses Integrated Nested Laplace Approximations [102].

The majority of the first half of this chapter has been published in *Medical Decision Making* as a paper entitled *A Review of Methods for the Analysis of the Expected Value of Information* [60]. The majority of the second half of this chapter has been published in *Statistics in Medicine* as a paper entitled *Estimating the Expected Value of Partial Perfect Information in Health Economic Evaluations using Integrated Nested Laplace Approximation* [59]. Additional elements have also been published in *Bayesian Cost Effectiveness Analysis with the* `R` *package* `BCEA` by *Springer* [5].

2.1 FINDING ANALYTICAL EXPRESSIONS FOR THE INNER INTEGRAL

The computational complexity in calculating the EVPPI by MC simulation is directly related to the inner conditional expectation in the first term of equation (4). Therefore, the computational effort required to calculate the EVPPI can be significantly reduced if the conditional expectation

$$\mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[\mathsf{NB}_t(\boldsymbol{\phi},\boldsymbol{\psi})\right]$$

can be expressed analytically as a function of $\boldsymbol{\phi}$. The PSA samples $\boldsymbol{\phi}_s$ for $s = 1, \ldots, S$ can then be inputted into this function to estimate this expectation without resorting to additional sampling.

One possible method for finding the analytical expression of this function, first introduced in [34], requires an approximately linear relationship between the *incremental* net benefit (INB) (the difference between the net benefit of two treatments, e.g. $\mathsf{INB} = \mathsf{NB}_1 - \mathsf{NB}_0$) and the parameters of interest $\boldsymbol{\phi}$. It also requires that $\boldsymbol{\phi}$ can be modelled (at least approximately) with a Normal distribution. If both these assumptions hold then it is possible to derive analytical

results based on the *unit loss integral* which then calculates the EVPPI in terms of the mean and variance of the conditional INB.

In this case, some additional sampling is required to determine the mean and variance of $\mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[\mathsf{INB}(\boldsymbol{\theta})\right]$. However, it is possible to estimate these values based on only two simulations from the conditional distribution, rather than $S >> 2$. This greatly reduces the computational time compared to the MC estimator but, as more recent methods rely solely on the PSA samples and have much looser assumptions, this method has rarely been used in practice.

A second analytic method was first developed in [46] and derives the analytic expectation under the assumptions that the net benefit is a linear function in the "nuisance" parameters $\boldsymbol{\psi}$ and that these are independent of the parameters of interest $\boldsymbol{\phi}$. In this case, the conditional distribution of $\boldsymbol{\psi} \mid \boldsymbol{\phi}$ is simply equal to the unconditional distribution. In addition to this, the net benefit can be written as follows:

$$\mathsf{NB}_t(\boldsymbol{\theta}) = \sum_{i=1}^{n_\psi} \psi^i f^i(\boldsymbol{\phi}),$$

where the superscript i indicates that ψ^i is the i-th element in the $\boldsymbol{\psi}$ vector and $f^i(\boldsymbol{\phi})$ is any known function of the parameters of interest. This means that the expectation can be written as

$$\mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[\sum_{i=1}^{n_\psi} \psi^i f^i(\boldsymbol{\phi})\right] = f^i(\boldsymbol{\phi}) \sum_{i=1}^{n_\psi} \mathsf{E}_{\boldsymbol{\psi}}\left[\psi^i\right],$$

so only the mean values for $\boldsymbol{\psi}$, from the initial PSA samples, are required to calculate the conditional expectation for each value of $\boldsymbol{\phi}$.

Therefore, provided the linearity assumption holds, or other expressions that would allow for the calculation of each expectation separately, such as a multiplicative term between two independent elements of $\boldsymbol{\psi}$ [12], it is possible to calculate the EVPPI using the PSA samples. However, this approximation deteriorates if the independence assumption is not valid [34]. More recently, Madan *et al.* [77] explored some re-parametrisations and approximations of the net benefit such that this method can be used when its original formulation does not conform to these conditions. For example, if the net benefit is non-linear in $\boldsymbol{\psi}$ but $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ are independent, it would then simply be possible to re-parametrise the problem so that NB_t is linear in a new parameter set $\boldsymbol{\psi}' = f(\boldsymbol{\psi})$.

They also explore the use of Taylor series to expand expressions that are non-linear in the parameters of interest. This is gives an approximation to the true EVPPI, albeit a sufficiently accurate one in many settings. Clearly, however, this method requires additional information about the model along with the computation of (possibly complex) Taylor series expansions making EVPPI calculation more challenging mathematically than computationally.

Finally, Oakley [86] developed a method where the incremental net benefit, as a function of $\boldsymbol{\phi}$, is emulated using a Gaussian Process (GP), see §2.6. Based on this GP, additional computations relying on quadrature methods are used to calculate the EVPPI. Due to the relative complexity of this method, it has never been applied in practice and has been entirely superseded by the method presented in §2.3 which also uses GPs.

A review paper [34] considered the performance of all these estimators (except the Gaussian Process method) and concluded that, provided the assumptions hold, all the methods discussed so far are suitable for calculating the EVPPI. After the publication of this paper, developments have been made so it is possible to calculate the EVPPI based solely on the PSA samples irrespective of model structure. As the PSA samples are already available in most health economic evaluations, these methods can be implemented with access to the underlying model or model assumptions.

2.2 APPROXIMATIONS FOR SINGLE PARAMETER EVPPI

The first developments calculated the EVPPI for a single parameter ϕ. In this setting, two methods were developed concurrently relying on the same underlying ideas: i) the first term for the EVPPI can be approximated by the average known-distribution net benefit value, provided the optimal decision remains constant; and ii) if a treatment option is optimal for one value of the parameter ϕ', it is still optimal for parameter values "close to" the point ϕ'. It is possible, therefore, to approximate the first term in (4) provided the net benefits can be partitioned such that there is an assurance that the optimal decision remains constant within each partition.

2.2.1 *The Strong and Oakley Method*

The first method [110] is based on the idea that if the parameter space is split into "small" subsets, the optimal decision is unlikely to change within each of these subsets, as the optimal decision is locally constant. Thus, to calculate the EVPPI, it is necessary to determine subsets of the simulated values of $\text{NB}_t(\boldsymbol{\theta})$ for which the simulated values of ϕ are similar. Practically, this is achieved by reordering the known-distribution net benefit values so that they have to same order as the simulated values for ϕ — note that since it is assumed that ϕ is a scalar, ordering is trivial. This list of ordered values is then split into M small sub-lists of length $L = \frac{S}{M}$. Within each sub-list the average known-distribution net benefit is calculated for each treatment option and the maximum within each subset is used as an estimate for the first term in equation (4), for each simulated value for ϕ.

Therefore, the following strategy can be used:

1. Obtain S PSA samples for $\boldsymbol{\theta}$ and $\text{NB}_t(\boldsymbol{\theta})$ for each treatment.

2. Sort the simulated values of the parameter of interest in ascending order — for simplicity, we write the sorted vector as $\varphi_1, \ldots, \varphi_S$, where $\min \phi = \varphi_1 < \ldots < \varphi_S = \max \phi$.

3. Re-order the estimated known-distribution net benefits as $\text{NB}_t(\varphi_1), \ldots, \text{NB}_t(\varphi_S)$, where $\text{NB}_t(\varphi_s)$ is the net benefit corresponding to the $s-$th ordered simulated value of ϕ.

4. Split the ordered list of known-distribution net benefits into $m = 1, \ldots, M$ sub-lists $\mathcal{L}_m$ of length L and compute the average in each of the sub-lists and for each treatment t to obtain
$$\widehat{\mathsf{NB}}_t^m(\boldsymbol{\theta}) = \frac{1}{L} \sum_{j \in \mathcal{L}_m} \mathsf{NB}_t(\varphi_j).$$

5. Compute $\widehat{\mathsf{EVPPI}} = \dfrac{1}{M} \displaystyle\sum_{m=1}^{M} \max_t \widehat{\mathsf{NB}}_t^m(\boldsymbol{\theta}) - \max_t \frac{1}{S} \sum_{s=1}^{S} \mathsf{NB}_t(\boldsymbol{\theta})$.

The approximation given by this method is highly sensitive to the value of M. If M is large, the formula for the EVPPI is similar to MC estimator the overall EVPI estimate, giving an upwardly biased estimate for the EVPPI. At the other end of the scale, if M is small, the EVPPI is approximately 0 giving a downwardly biased estimate. Strong and Oakley [110] use a normal approximation to estimate the upward bias of the EVPPI calculated using this method. They suggest choosing M to be the largest number of subsets such that the upward bias falls below a pre-specified threshold value. In this manner, the upward bias of the EVPPI estimate, present when the chosen value of M is too large, is controlled while the downward bias, present for small values of M, is likely to be avoided.

Practically, this bias calculation drives the estimate of the EVPPI. Specifically, if a parameter has a high EVPPI then the bias is minimised with a higher value of M than with smaller values for the EVPPI. This means that the bias must be estimated separately for each EVPPI calculation to ensure consistency. This bias estimation method assumes a multivariate normal distribution of the conditional expectation value within each sub-list $\mathcal{L}_m$. Therefore, there is an implicit assumption in the calculation the EVPPI using this method that the normal approximation is "good enough" to approximate the true bias.

2.2.2 *The Sadatsafavi et al. Method*

This second method for single parameter EVPPI [104] can be thought of as an extension to the previous one, although both were developed concurrently. In this case, the ordered list

$$\mathbf{NB}(\boldsymbol{\theta}) = \begin{pmatrix} \mathsf{NB}_0(\varphi_1) & \cdots & \mathsf{NB}_0(\varphi_S) \\ \mathsf{NB}_1(\varphi_1) & \cdots & \mathsf{NB}_1(\varphi_S) \\ \cdot & \cdot & \cdot \\ \mathsf{NB}_T(\varphi_1) & \cdots & \mathsf{NB}_T(\varphi_S) \end{pmatrix}$$

is split in an informed manner. The algorithm searches for the point(s) at which the parameter of interest is directly responsible for a change in the optimal decision. Although, in general, variations in a parameter may induce a change in the optimal decision a large number M of times, in practice a single parameter is unlikely to modify it more than once or twice over the range of values selected for the willingness-to-pay. Provided we search for the correct number of decision changes for the parameter in question, the EVPPI estimator found using this method tends to the true EVPPI value as the number of PSA samples goes to infinity.

The algorithm developed by Sadatsafavi *et al.* is identical, in steps 1-3, to that of Strong and Oakley described in the previous section, but then proceeds as follows:

4. Split the ordered list of net benefits into $m = 1, \ldots, M$ sub-lists $\mathcal{L}_m$. In this case, the list of known-distribution net benefit values is split at M points, $\left(\varphi^{\mathcal{L}_1}, \ldots, \varphi^{\mathcal{L}_M}\right)$, known as the *segmentation vector*. This means that the $m-$th sub-list contains all the net benefit values calculated with the value φ such that $\varphi^{\mathcal{L}_m} \leq \varphi < \varphi^{\mathcal{L}_{m+1}}$.

5. Calculate and store $\widehat{\mathsf{NB}}_t^m(\boldsymbol{\theta}) = \dfrac{1}{L} \sum\limits_{j \in \mathcal{L}_m} \mathsf{NB}_t(\varphi_j)$.

6. Maximise over all possible segmentation vectors to compute

$$\widehat{\mathsf{EVPPI}} = \max_{(\varphi^{\mathcal{L}_1}, \ldots, \varphi^{\mathcal{L}_M})} \frac{1}{M} \sum_{m=1}^{M} \max_t \widehat{\mathsf{NB}}_t^m(\boldsymbol{\theta}) - \max_t \frac{1}{S} \sum_{s=1}^{S} \mathsf{NB}_t(\boldsymbol{\theta}).$$

As the properties of the EVPPI estimate are dependent on searching for the correct number of decisions M, Sadatsafavi *et al.* suggest a systematic method for choosing M, based on a visual tool for determining the number of decision changes. This tool plots the cumulative sum of the differences between the known-distribution net benefits for the two treatment options, ordered in the parameter of interest ϕ.

As a function of ϕ we define

$$\hat{C}_{t,t'}(\varphi) = \frac{1}{S} \sum_{s=1}^{S} I(\phi_s < \varphi) \left[\mathsf{NB}_t(\phi_s) - \mathsf{NB}_{t'}(\phi_s)\right]. \tag{9}$$

The extrema in this estimator will correspond to a change in optimal decision. For example, if t is the optimal decision until $\phi = 1$ then as ϕ increases from 0 to 1, only positive terms will be summed, then as the optimal decision changes after 1, negative terms will be summed and therefore the value of $\hat{C}$ will decrease.

We then plot $\hat{C}$ against the parameter of interest and visually search for maxima or minima. In some cases, the visual tool can be used easily (Figure 3 LHS), but in others it can be challenging to determine whether an extremum exists (Figure 3 RHS). This adds complexity to the method as correctly identifying the number of decision changes is vital. This is seen most clearly when the EVPPI is 0, i.e. when the parameter ϕ does not affect the optimal decision. In this case $M = 0$, however, choosing $M = 1$ will estimate an EVPPI which is strictly positive and therefore upwardly biased.

The advantages of these methods are that neither rely on additional sampling from the health-economic model once the PSA samples are available. However, both are very dependent on inputs that can be challenging to determine, or rely on normal approximation. Therefore, an alternative method has been developed to calculate EVPPI using the PSA samples. This method can also be used for multi-parameter EVPPI.

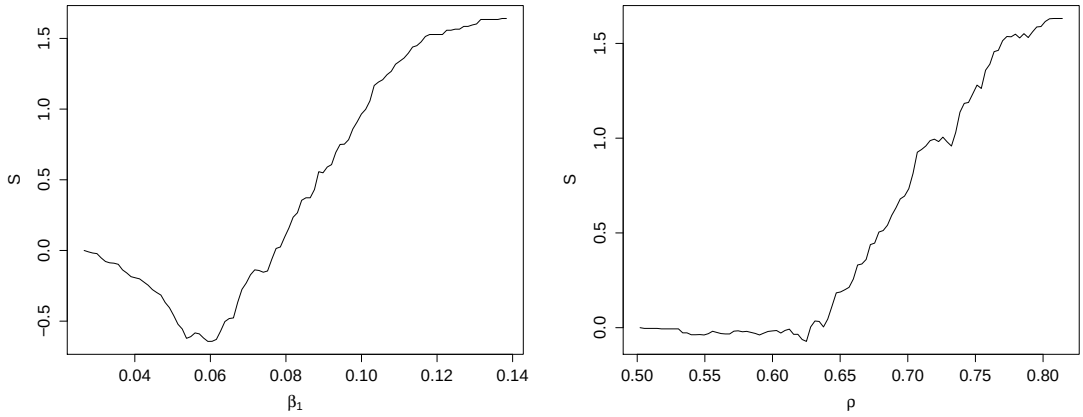


Figure 3: An example of two plots showing the use of the visual tool for the Sadatsafavi *et al.* for calculating single parameter EVPPI.

2.3 NON-PARAMETRIC REGRESSION FOR EVPPI CALCULATIONS

The unit loss integral method [34] and the GP method [86] are both based on the idea that the incremental net benefit can be approximated using regression (linear or non-parametric). These regression formulæ can then simplify the EVPPI calculations. Strong et al. [111] develop this idea further and demonstrate that the conditional expectation of the known-distribution net benefit can be approximated using regression methods and based directly on the PSA samples.

In particular, since the conditional expectation will not have a known functional form, flexible regression methods should be used. These methods assume that the conditional expectation of the known-distribution net benefit is a smooth function of the important parameters $\boldsymbol{\phi}$ but make no other assumptions. This leads to a general method for calculating the EVPPI in multi-parameter settings. For each simulated value of the parameters s, Strong et al. propose to approximate the inner conditional expectation in (4) as

$$\mathrm{NB}_t(\boldsymbol{\theta}_s) = \mathrm{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}_s}\left[\mathrm{NB}_t(\boldsymbol{\phi}_s, \boldsymbol{\psi})\right] + \varepsilon_s,$$

with $\varepsilon_s \sim \mathrm{Normal}(0, \sigma_\varepsilon^2)$. Furthermore, the conditional expectation can be thought of as a function of $\boldsymbol{\phi}$ only, as the *conditional* expectation is only dependent on the value of $\boldsymbol{\phi}$. Therefore, the problem can be formulated as

$$\mathrm{NB}_t(\boldsymbol{\theta}_s) = g_t(\boldsymbol{\phi}_s) + \varepsilon_s,$$

where $g_t(.)$ is an unknown function of $\boldsymbol{\phi}$ that is estimated using non-parametric regression.

Strong *et al.* [111] suggest two alternative non-parametric regression methods: Generalised Additive Models (GAMs) [54] and Gaussian Processes (GPs) [98]. GPs are a flexible regression method based on the multivariate normal distribution. GPs introduce flexibility by modelling the covariance matrix of the net benefits dependent on the $\boldsymbol{\phi}$ values and some

hyperparameters that are estimated using the net benefits. This hyperparameter estimation, formally introduced in §2.6.1, can be computationally costly for larger numbers of net benefit observations c.f. §2.6.

GAMs model the net benefit as a sum of "smooth" functions of the parameters $\boldsymbol{\phi}$. In standard software, splines [41, 124] are used to estimate these smooth functions. Broadly splines are a class of flexible functions defined using piecewise parametric functions which are then continuously joined together at specific points known as knots. The use of splines implies that a large number of "regression" parameters [125] must be estimated to fit a GAM. This implies that GAMs cannot be fitted for "large" $\boldsymbol{\phi}$ as the number of required regression parameters exceeds the number of data points. Both these non-parametric regression methods offer a large amount of flexibility but other model specifications could be used.

In a general sense, the EVPPI estimate is calculated by means of the following algorithm:

1. Obtain S PSA samples for $\boldsymbol{\theta}$ and $\text{NB}_t(\boldsymbol{\theta})$ for each treatment.

2. For each treatment option t, fit a regression curve with $\boldsymbol{\phi}_s$ as the observed "covariates" and $\text{NB}_t(\boldsymbol{\theta}_s)$ as the observed "response".

3. For each treatment t, find the $s-$th fitted values $g_t(\boldsymbol{\phi}_s)$ by inputting the observed values $\boldsymbol{\phi}_s$ into the regression curves.

4. Calculate $\widehat{\text{EVPPI}} = \frac{1}{S}\sum_{s=1}^{S} \max_t g_t(\boldsymbol{\phi}_s) - \max_t \frac{1}{S}\sum_{s=1}^{S} g_t(\boldsymbol{\phi}_s)$.

Theoretically, these methods are amongst the most complex for calculating the EVPPI. However, due to the flexibility of non-parametric regression, "black-box" calculations that require little input from the user are possible. Specifically, the `R` package `mgcv` [126] for GAM regression and specific priors for the hyperparameters of the GP cf. §2.6 allows function to be created to automate this process [108]. This has also led to the creation of web applications such as *Sheffield Accelerated Value of Information* (`SAVI` [109]) and `BCEAWeb` [5] which allows users to calculate EVPPI with using `R`.

2.3.1 *GAM regression within `mgcv`*

To understand the assumptions made by these "black-box" functions, we briefly present how GAMs are fitted within the `mgcv` package. For this analysis, EVPPI calculations were undertaken using the default settings within `mgcv` package. Therefore, we now briefly explore the assumptions made for these defaults. Firstly, for a uni-variate ϕ, fitting a GAM implies that $g_t(\phi)$ is estimated as a weighted sum of basis functions

$$g_t(\phi_s) = \sum_{k=1}^{K} \alpha_k B_k(\phi_s).$$

To fit the model, it is necessary to determine or estimate K, α_k and $B_k(\phi_s)$. Within `mgcv`, models for uni-variate predictors are specified using the `s()` function which uses thin plate

regression splines as a basis function [124]. The number of knots for this class of splines is given deterministically from the dimension of the spline basis, K.

K can be chosen, conditional on the data, within `mcgv` to ensure optimal properties. Specifically, K is chosen to *overfit* to the data being modelled. This ensures that the GAM is sufficiently flexible to capture the true relationship between ϕ and $g_t(\phi)$. Then, to prevent this overfitting, the regression coefficients α_k and the coefficients of the basis splines are penalised using a quadratic penalty. This ensures that deviations from smoothness are supported by the data rather than encouraged by the model specification. As with all penalisation methods, the strength of the penalty must be chosen before estimating the model parameters. Within `mgcv`, however, this penalisation strength is chosen automatically using a method presented in [125].

For multivariate $\boldsymbol{\phi}$, Strong et al. [111] suggest using a tensor product form of GAM regression to model potential interactions between the parameters in $\boldsymbol{\phi}$. Within `mgcv`, this is specified using the `te()` function. By default, this model specification uses cubic regression splines as its basis [41], where K and the penalty for the penalised regression are, by default, selected to ensure accurate model fit.

2.4 COMPARISON OF METHODS

As the performance of the EVPPI estimation methods has not been compared since the Coyle and Oakley review [34], we assess the performance of these more recent methods by comparing them with EVPPI estimates obtained using MC simulations. We begin by presenting the two case studies used to compare the methodologies. We then discuss how the analysis was performed before presenting the results of our analysis. We also compare the standard error, computation time and ease of application before concluding that the Sadatsafavi *et al.* method is only suitable for 2 decision settings and that the non-parametric regression methods offer the best option for calculating the EVPPI.

2.4.1 *Case Studies*

2.4.1.1 *Influenza Vaccine*

Baio and Dawid [7] consider a Bayesian health-economic model synthesising information from several sources is proposed to analyse the effect of the influenza vaccine on health-outcomes and costs. This example is relatively simple but complex enough for the posterior distribution to be intractable analytically. Therefore, the parameters values must be sampled from their joint posterior distribution using MCMC methods. Updating the joint posterior may induce correlations for parameters that are independent a priori, meaning that the independence structure of the parameters is not known.

Two treatment options are considered, either the vaccine is available to the population ($t = 1$), or not ($t = 0$). If a patient gets influenza, they are treated with anti-viral drugs and will often visit the GP. Complications may occur, including pneumonia and hospitalisation – in which

case there will be some indirect costs such as time off work. The cost of the treatment is the acquisition cost of the drugs, the time in hospital, the GP visits and the cost of the vaccine. The benefit of the treatment is measured in QALYs (Quality Adjusted Life Years), where each adverse effect contributes negatively to the benefit.

There are 28 key parameters in the model representing the probability of infection, the reduction in risk due to the vaccine, the occurrence of complications, the monetary costs of the interventions and the QALY loss due to different health states. A detailed discussion of the full model parametrisation is presented in [7], based on [33]. The single parameter EVPPI was calculated for all of these 28 parameters using all the methods outlined.

2.4.1.2 *FEAST Trial*

This example pertains to the FEAST trial for fluid resuscitation in the treatment of malaria [77]. This is a decision tree model where both the probabilities of progression as well as the QALYs and costs are estimated from various sources. The parameters in this setting are either estimated using Bayesian updating based on a set of randomised controlled trials or sampled directly from an informed prior distribution. Therefore, in this example, the independence structure of some of the parameters is known. The example has four potential treatment options, 3 different fluids and a standard of care arm.

The use of fluids is aimed at reducing the number of patients living with long term neurological sequelae (NS) along with the risk of death. Living with long term health complications has an added monetary and QALY cost. This model has 9 parameters and is used to test the method proposed by Madan *et al.* [77]. Therefore, the Monte Carlo EVPPI values used in this analysis as the "truth" are taken directly from that paper.

For each active interventions $i = 1, 2, 3$, the parameters in this case study represent the effect of the different fluids on mortality: d_i^M; and on the NS risk: d^S (which is assumed constant across the fluids); the log-odds of death without fluids: α; the probability of NS at 28 days: p^B and 6 months: p^L; and the QALY loss per fatality: q^M and per case of NS: q^S.

2.4.2 *Analysis*

2.4.2.1 *Influenza Vaccine*

For this case study, only the single parameter EVPPI is calculated. As previously discussed, both estimation methods only suitable for single parameter estimation have input values that greatly affect the estimate. Therefore, to determine these values for the Sadatsafavi *et al.* method, the suggested visual tool was used for each of the 28 parameters. It was decided that 23 out of the 28 parameters change the optimal decision once as one extremum was identified, the other parameters have no effect – these correspond to an EVPPI equal to 0 in Table 4.

For the Strong and Oakley [110] method, the upward bias of the EVPPI estimate was assessed for all the parameters in the data set. The number of subsets was taken as the largest

number of subsets such that the upward bias remains below 0.1. The magnitude of the bias differed for each parameter, and therefore the number of subsets needed to reduce the upward bias below 0.1 was between 20 and 2. A larger number of subsets is associated with a larger estimate for the EVPPI and so clearly this procedure impacts on the EVPPI estimate.

For the two-step MC procedure the outer loop had 1000 observed posterior samples for all the parameters. As the case study is a Bayesian model with analytically an intractable posterior distribution, the inner loop of the two-step MC procedure used MCMC methods to draw a sample from this joint posterior distribution for all the parameters. To ensure convergence and reduce autocorrelation, we took 2 chains of $100,000$ samples each for the inner loop, with a burn-in of $9,500$ and thinning of 500. Therefore, the eventual sample of the known-distribution net benefits had 1000 observations. A crude estimate of the MC error for these estimates gives the error as around 10-30% of the underlying EVPPI value. This standard error is estimated by taking the variance of the maximum between 0 and the conditional expectation conditional on each parameter, before dividing by 1000, the size of the outer loop, and adding Monte Carlo error related to estimating the expected known-distribution net benefit under current information. This gives the standard error that arises from the outer loop expectation but entirely disregards any error coming from the inner loop. Ideally, the standard error would be estimated following the procedure set out in [12] but simulations were not available for this analysis.

Note, therefore, that the point estimates given in Figure 4 are subject to substantial MC error. Ideally we would have reduced this error by using a greater number of simulations in both the inner and outer loop. This was, however, computationally infeasible. Additionally, the nested MC simulation for the EVPPI is subject to an upward bias [12] which will impact the comparison with the alternative methods.

For the nested MC estimator, we require an estimate of the value of the optimal treatment under current information. To calculate this, we sampled from our MCMC procedure $50,000,000$ times, with a burn-in of $9,500$ and then thinned to $500,000$. This gives a PSA sample for the known-distribution incremental net benefit with a mean of 0.041. This means that the value of the optimal treatment under current information is 0.041. The MC error on this mean of the known-distribution incremental net benefit is estimated as 0.017 by taking the variance of the sample and dividing by the sample size. This assumes that the MCMC samples are independent but this is a reasonable assumption given that the sample has been thinned.

Finally, the two different non-parametric regression methods suggested by Strong *et al.* [111] were used to calculate the EVPPI by approximating the inner expectation in (4). This analysis was carried out using the functions in [108]. Within this code, the default single parameter GAM regression c.f. §2.3.1 is used and for GPs, Strong et al. use a specific model structure to allow for the use of analytical results, therefore, no model structure needs to be specified by the user and functions are provided to calculate the EVPPI with no additional inputs, with the default 500 PSA samples used to estimate the hyperparameters.

2.4.2.2 *FEAST Trial*

This analysis compares the performance of the different EVPPI estimation methods for all 12 subsets considered in Madan *et al.* [77]. Of these sets, 3 contain more than one parameter. Therefore, this example allows the comparison of the multi-parameter EVPPI calculation methods as well as methods for single parameter EVPPI in a multi-decision setting. All the analysis performed in this section was performed on 5 million samples from the FEAST analysis performed by Madan *et al.*. This was thinned to 1000 as there was substantial autocorrelation present in the sample and to reduce computational times.

All methods must be modified slightly for multi-decision, multi-parameter problems. For the single parameter methods, the input values must be determined. To use the Sadatsafavi *et al.* method for multi-decision problems it is necessary to decide the number of decision changes for each pair of decisions. Therefore, in this setting, we used the visual tool for 6 different pairs of decisions for each calculation. This decreases the accuracy of this method as the visual tool can be difficult to interpret and with 6 decisions per EVPPI calculation, there is a higher chance of an error. In this case, the number of decision changes varied between 0 and 3.

For the Strong and Oakley [110] method the bias reduction technique was used. In this setting the number of blocks was relatively small for all EVPPI calculations, varying only between 5 and 2. These values were determined by insisting that the upward bias for the EVPPI standardised by the EVPI had to fall below 0.02. This rescaling was used because the absolute size of the bias is directly proportional to the absolute size of the values we are estimation and the EVPI in this case was 541.

Extending the non-parametric regression methods to multi-decision models involves fitting an extra regression curve per *incremental* decision. Therefore, 3 regression curves must be found for each EVPPI calculation. As the GAM regression using the tensor product becomes infeasible for "large" parameter subsets, the GP regression is the only option for two of the three multi-parameter EVPPI calculations. Again, we use the default options of 500 PSA samples to estimate the hyperparameters for the GP regression. For the multi-parameter GAM method, we model full interactions.

All the analysis were run using `R` (version $3.2.1$). The MCMC simulations were run using `JAGS` [93] which was called using the `R` package `rjags` [92]. Computations were performed on a PC with an Intel i7 processor and 8 GB of RAM.

2.4.3 *Comparison of current EVPPI estimation methods*

Figure 4 displays the single parameter EVPPI estimates for the five different methods for the Vaccine Example. All the methods give approximately the same results for each parameter. The two parameters with the highest EVPPI are the same across all five methods: β_1 and ω_1, representing respectively the probability and reduction in QALYs due to an influenza infection. Although there is some discrepancy in the ordering, the results only differ slightly, certainly

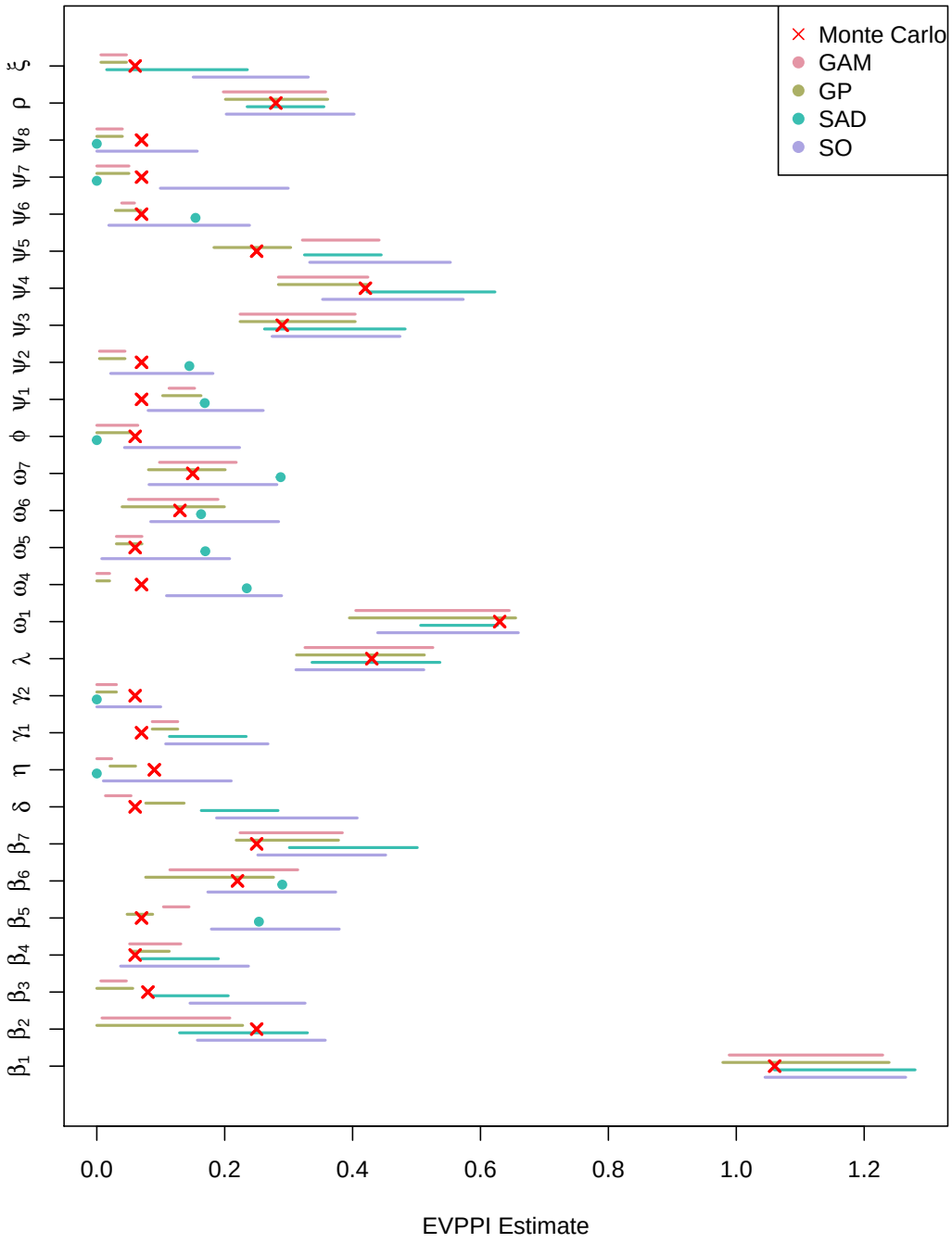


Figure 4: The EVPPI values, plus or minus one standard error for the parameters in the Vaccine example using 5 different methods - Strong and Oakley single parameter estimation (purple - SO), Sadatsafavi *et al.* single parameter estimation (blue - SAD), Gaussian Process (green - GP) and GAM regression (pink - GAM) and two-step Monte Carlo simulation (red crosses - MC)

in comparison to the Monte Carlo error of the value of the optimal treatment under current information which is 0.02.

Figure 5 gives the EVPPI for all the parameter sets considered for the FEAST example. For this example, the Sadatsafavi *et al.* method struggles to retrieve the true EVPPI. Therefore, this method is not recommended in the multi-decision setting due to the difficulty in ascertaining

the correct number of decision changes. The other methods give similar answers across all the different parameter sets, with some difficulties with the Strong and Oakley method and the GP regression for d_3^M and d_4^M (representing the effect of the saline and gelofusine fluids on mortality, respectively) and for the GAM regression for the parameter subset of size four, where the estimate is more than one standard error away from the MC "truth". This suggests that the GAM regression may be preferred in terms of accuracy for single parameter EVPPI whereas the GP performs better in settings with 4 or more parameters.

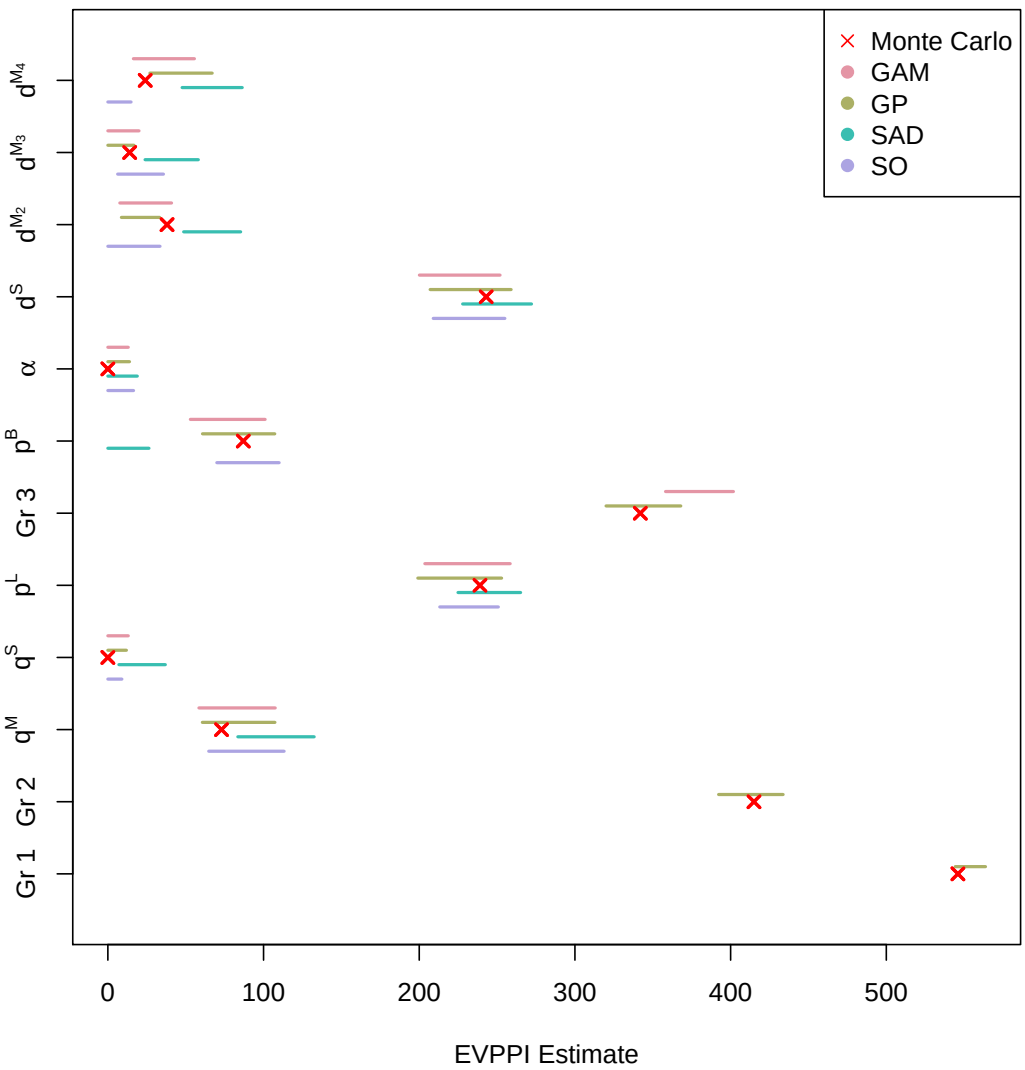


Figure 5: The EVPPI values, plus or minus one standard error for the parameters in the FEAST example using 5 different methods - Strong and Oakley single parameter estimation (purple - SO), Sadatsafavi *et al.* single parameter estimation (blue - SAD), Gaussian Process (green - GP) and GAM regression (pink - GAM) and two-step Monte Carlo simulation (red crosses - MC). Groups 1, 2 and 3 contain the following parameters respectively, (d_2^M, d_3^M, d_4^M, α, p^B, d^S, p^L), (d_2^M, d_3^M, d_4^M, α, p^B, d^S, q^M) and (d_2^M, d_3^M, d_4^M, d^S)

Broadly speaking, these results show that the approximation methods offer similar results for health-economic models with only 2 decision options and so can all be used to estimate

the EVPPI. The Sadatsafavi *et al.* method should be reserved for this setting. It is important to note that the accuracy of the estimate of the EVPPI should only be reported to 2 or 3 significant figures, even when the analysis gives a much greater accuracy of calculation as all these methods only approximate the true EVPPI.

Given that, in general, the methods produce broadly similar estimates for the EVPPI, it seems pertinent to investigate the computation time required to produce the estimate. For the Vaccine example, the single-parameter estimation methods and the GAM regression took less than 0.01 seconds to find single parameter EVPPI estimates. In contrast, the GP regression takes around 2-4 seconds, which is still fast in comparison to the MC estimation procedure that takes around 55 minutes to calculate a single parameter EVPPI estimate. Therefore, all the approximation methods offer significant computational savings in the dual-decision setting, with the GP regression taking noticeably longer.

For the FEAST example, the single parameter estimation times are similar, although the GP method increased for this example to 7-9 seconds as hyperparameters need to be estimated for each (incremental) decision. The multi-parameter estimation times are significantly longer. In the four parameter example, the GP method took 85 seconds, while the GAM regression method took 184 seconds. Madan *et al.* [77] give the one step MC estimation time as 84 seconds while the two step MC is 31 minutes. Consequently, with a view to obtaining a general purpose method, the GP regression is preferred for four parameter estimation in light of the computation time and the lower accuracy of the GAM estimation method. Additionally, the Madan *et al.* method is recommended in situations where it can be employed.

For the seven parameter examples, the GP regression times are 229 and 279 seconds respectively. The Madan *et al.* estimation times are given as 9 and 10 seconds for single step and 26 minutes for two step Monte Carlo. Clearly, therefore, in terms of computation time the Madan *et al.* method is superior. However, it must be stressed that this method relies on inspection of the net benefit function along with algebraic manipulation. Therefore, while the computation time is low the time spent analysing the net benefit function will clearly be greater than 279 seconds.

In addition to the computational time, it is useful to investigate the standard error for each of the estimation procedures. This error is represented in Figures 4 and 5 and is fairly consistent across the different methods and across the different parameter subsets sizes. In general, larger estimates of the EVPPI are associated with a relatively smaller standard error.

The standard errors were calculated using methods specific to each EVPPI estimation method. Firstly, for the GP regression method, the standard error is calculated by repeated resampling from a multivariate normal distribution. Specifically, the GP regression method specifies a multivariate normal model for the fitted values $g_t(\boldsymbol{\phi}_s)$. To calculate the standard error, simulations are taken from this multivariate normal distribution and each of these are used to calculate the EVPPI. The standard error for the GP method is then estimated as the standard deviation of these EVPPI values. In this sense, the GP standard error estimate is entirely dependent on the estimated model parameters and does not reflect additional uncertainty that would arise from the model fitting procedure.

For the GAM regression method, the standard error is calculated using a method similar to the GP regression estimates. However, as the GAM method does not define a multivariate normal distribution for the fitted values, this distribution is estimated before calculating the standard error. Firstly, the `mgcv` package outputs the values of the basis functions for each $\boldsymbol{\phi}_s$ and the estimated values of the coefficients α_k. These estimates can be combined to give the mean vector for the fitted values using the formula in §2.3.1. The `mgcv` package also estimates the covariance matrix of the regression parameters α_k which can then be used to give the covariance matrix of the fitted values. This gives a mean vector and covariance matrix for a multivariate normal approximation to the fitted values. From this distribution, the standard errors for GAM regression are calculated in the same way as for the GP regression. Therefore, the GAM standard error is estimated conditional on the model and it assumes that the basis functions accurately represent the underlying fitted values.

For the Strong and Oakley single parameter method, the standard error is estimated by calculating the within-block variance and then averaging over all the blocks. More specifically, the standard error averages over the within-block variance for the treatment that is considered optimal within that block. This method gives the variance of the first term of the EVPPI estimator which must then be added to the variance of the value of the optimal treatment under current information. This standard error calculation has parallels with the optimal method for calculating the standard error of the nested MC estimator.

Finally, as no method is recommended to calculate the standard error of the Sadatsafavi et al. method, we used bootstrapping to resample the incremental net benefit values before repeatedly calculating the EVPPI. The standard error was estimated as the standard deviation of these bootstrapped estimates. In some settings, estimated the standard error as 0 implying that bootstrapping is unlikely to accurately estimate the standard error of this method.

2.4.4 *Computational Time*

To investigate the computational time required for the multi-parameter EVPPI estimation methods further, the EVPPI was calculated for increasing subsets for the Vaccine example (accuracy was not tested). The computational time was taken for subsets containing between 1 and 16 parameters for the GP regression method and 1 and 4 parameters for the GAM regression as the number of parameters exceeds the number of data points for larger subsets.

The timing for the GAM regression is 0.13, 0.11, 0.66 and 122 seconds for increasing subsets. Therefore, in terms of computational time the GAM regression is recommended for subsets with fewer than 4 parameters. However, for four or more parameters, computationally, it is advisable to use the GP regression. In addition to this, the GAM regression appears to be less accurate for larger subsets as seen with the FEAST example. This is due to overfitting as the number of parameters estimated to fit the GAM approaches the number of data points.

Figure 6 shows the computational time required to calculate the GP estimate for increasing parameter subset sizes. The more parameters in $\boldsymbol{\phi}$ the greater the computational time. Additionally, we note that the computational time increases exponentially with the number of

parameters. This causes difficulties when large sample sizes are coupled with larger parameter subsets as the cost of calculating the EVPPI using the GP method is proportional to S^3, where S is the PSA sample size, [111], along with a high cost associated with the number of parameters.

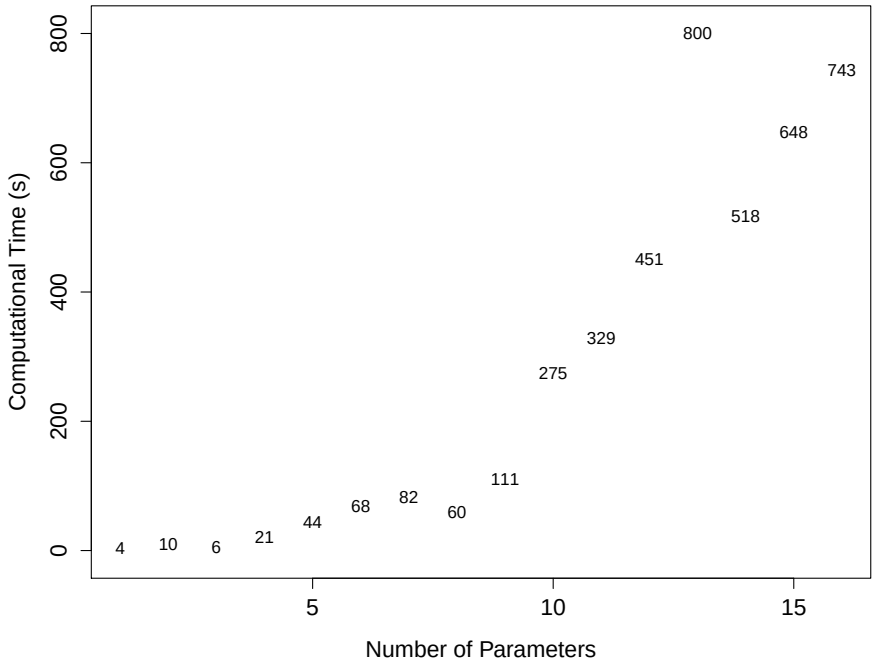


Figure 6: The computational time required for EVPPI estimation for the GP regression method, varying with the number of parameters for which the EVPPI is being calculated.

2.5 SUMMARY OF EVPPI CALCULATION METHODS

Thus far, we have presented the currently available methods for approximating the EVPPI along with a formal comparison of the 4 more recent methods. One of these methods, the Sadatsafavi *et al.* method, is unsuitable for multi-decision problems due to difficulties in ascertaining the inputs required for accurate estimation using this method. The other 3 methods gave broadly consistent results when compared with MC estimator. All the methods also have similar standard error, meaning that they should be compared on their computational time and ease of implementation.

One of the remaining three methods can only be used for single parameter EVPPI which limits its applicability. Additionally, a thorough understanding of the bias reduction tool for identifying the correct input value is essential. Although standard functions are available for calculation using this method [6, 110], the input values are still needed and a naive implementation of this function will lead the experimenter to grossly over/under-estimate the single parameter EVPPI. Therefore, this method is largely superseded by the non-parametric regression methods.

As the GAM regression requires negligible computational effort for single-parameter EVPPI, it is the best method under these circumstances. In low-dimensional settings, the GAM method is accurate as well as computationally efficient and therefore there is little merit in pursuing an alternative EVPPI calculation method when considering a small number of parameters of interest. However, for larger subsets, the computational time for this method increases substantially as well as losing accuracy. Additionally, GAM regression is mathematically unstable for larger subsets of parameters (≥ 5) [111] and so cannot be used to approximate the EVPPI.

For larger parameter subsets, ≥ 4, Gaussian Process regression is more accurate than GAM regression and can be used in settings where this method breaks down. Unfortunately the Gaussian Process method takes substantially more computational time as the number of parameters in $\boldsymbol{\phi}$ increases. In addition to this, the computational cost for GP regression method is proportional to S^3, so the GP becomes prohibitively expensive when large number of parameters of interest $\boldsymbol{\phi}$ are coupled with a large PSA simulation size.

Therefore, we propose a method that allows for fast computation of the EVPPI in these high dimensional settings. This method is based on extension to the standard GP regression approach exploiting results from the spatial statistics literature.

2.6 FAST GAUSSIAN PROCESS REGRESSION

We have seen that GP regression can be used to estimate the EVPPI correctly but that the computational time for the EVPPI estimate can be prohibitive in high dimensions. Therefore, we have developed a fast approximation to the GP regression method that significantly reduces the computation time whilst retaining the accuracy in many settings. To present this method, we formally introduce GP regression, focusing initially on the GP regression method used by Strong *et al.*.

Gaussian Processes are a family of stochastic processes used in statistics and machine learning for non-parametric regression, classification and prediction [37, 98] and can be thought of as an extension of the multivariate Normal distribution to an infinite vector of observations [42, 98]. Strictly speaking, a GP is an infinite collection of random variables, any subset of which follows a multivariate Gaussian distribution [97]. A GP is entirely defined in terms of its mean and covariance *functions* [53, 76], which calculate the mean vector and covariance matrix for each subset of random variables depending on some input values and a small set of hyperparameters. These inputs determine the specific mean and variance for each random variable in the process. Consequently, GPs can be used for regressing random variables on a set of input values.

To fit a GP for non-parametric regression, the general form of the mean and covariance function is specified by the modeller. In general, the covariance function is a taken as a decreasing function of the “distance” between any two input values, i.e. points that are “closer” have a higher correlation [88, 98] where “distance” and “closeness” can be measured using any metric. These functions typically depend on a set of hyperparameters; for example, the

covariance function is often defined in terms of a smoothness parameter that determines the similarity between two points "close" together and a GP marginal variance parameter. Once these general functions are specified, problem-specific values for the hyperparameters must be determined.

In a Bayesian setting, vague and conjugate priors have been proposed for the hyperparameters allowing for partially analytically tractable posterior distributions [89, 111]. Integration and numerical optimisation can then be used to find maximum *a posteriori* estimates of the GP parameters. Each numerical optimisation step requires the inversion of a dense matrix which has a computational cost of S^3, as quoted in §2.5. However, this computational cost is often small in comparison to nested MC simulation.

Recall that the non-parametric regression method [111] requires estimating the function $g_t(\boldsymbol{\phi})$ using the model structure specified as

$$\text{NB}_t(\boldsymbol{\theta}_s) = g_t(\boldsymbol{\phi}_s) + \varepsilon_s. \tag{10}$$

with $\varepsilon_s \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma_\varepsilon^2)$. Assuming a GP structure for the functions $g_t(\cdot)$ in a linear regression framework effectively amounts to modelling

$$\begin{pmatrix} g_t(\boldsymbol{\phi}_1) \\ g_t(\boldsymbol{\phi}_2) \\ \vdots \\ g_t(\boldsymbol{\phi}_S) \end{pmatrix} \sim \text{Normal}(\boldsymbol{H\beta}, \boldsymbol{\Sigma}), \tag{11}$$

where: $\boldsymbol{H}$ is a design matrix;

$$\boldsymbol{H} = \begin{pmatrix} 1 & \phi_1^1 & \cdots & \phi_1^P \\ 1 & \phi_2^1 & \cdots & \phi_2^P \\ \vdots & & \ddots & \\ 1 & \phi_S^1 & \cdots & \phi_S^P \end{pmatrix}; \tag{12}$$

$\boldsymbol{\beta}$ is the vector of regression coefficients describing the linear relationship between the parameters of interest $\boldsymbol{\phi}$ and the conditional expectation of the net benefits; and the covariance matrix $\boldsymbol{\Sigma}$ is determined by the covariance function $\boldsymbol{\mathcal{C}}$, a matrix operator whose elements $\mathcal{C}(r,s)$ describe the covariance between any two points $g_t(\boldsymbol{\phi}_r)$ and $g_t(\boldsymbol{\phi}_s)$.

Strong et al. [111] use a squared exponential, also known as an exponentiated quadratic, covariance function $\mathcal{C}_{\text{Exp}}$, defined by

$$\mathcal{C}_{\text{Exp}}(r,s) = \sigma^2 \exp\left[\sum_{p=1}^{P}\left(\frac{\phi_r^p - \phi_s^p}{\delta^p}\right)^2\right] \tag{13}$$

where ϕ_r^p and ϕ_s^p are the r-th and the s-th simulated value of the p-th parameter in $\boldsymbol{\phi}$, respectively. For this covariance function, σ^2 is the GP marginal variance and δ^p defines the smoothness of the relationship between two values that are "close together" in dimension p.

For high values of δ^p the correlation between the two conditional expectations with similar values for ϕ^p is small. The δ^p values are also treated as hyperparameters to be estimated from the data.

Combining equations (10) and (11), we can directly model the "observed" vector of net benefits as

$$\begin{pmatrix} \text{NB}_t(\boldsymbol{\theta}_1) \\ \text{NB}_t(\boldsymbol{\theta}_2) \\ \vdots \\ \text{NB}_t(\boldsymbol{\theta}_S) \end{pmatrix} \sim \text{Normal}(\boldsymbol{H\beta}, \boldsymbol{C}_{\text{Exp}} + \sigma_\varepsilon^2 \boldsymbol{I}). \tag{14}$$

The model in (14) includes $2P+3$ hyperparameters: the $P+1$ regression coefficients $\boldsymbol{\beta}$, the P smoothness parameters $\boldsymbol{\delta} = (\delta^1, \ldots, \delta^P)$, the marginal variance of the GP σ^2 and the residual error σ_ε^2, also known as "nugget variance". In this sense, the simulated output used to perform PSA $\boldsymbol{\phi}_s$ become covariates used to regress onto the known-distribution net benefits. Clearly, therefore, the number of parame ters in $\boldsymbol{\phi}$ affect the EVPPI computational time, as seen in §2.4.4.

2.6.1 *Hyperparameter estimation for GP regression*

To determine hyperparameters for the GP, Strong et al. [111] set the following priors for the model parameters:

- Each regression parameter in $\boldsymbol{\beta}$ is given an independent vague uniform prior $\pi(\beta^p) \propto 1$ for $p = 1, \ldots, P+1$.
- The GP marginal variance σ^2 is given a weak inverse gamma prior $\sigma^2 \sim IG(0.001, 0.001)$.
- The nugget variance σ_ε^2 is given a weak inverse gamma prior $\sigma_\varepsilon^2 \sim IG(0.001, 1)$.
- The smoothness parameters $\boldsymbol{\delta}$ are each assumed to have a weak normal prior, $\boldsymbol{\delta}^p \sim N(0, 10^6)$ for $p = 1, \ldots, P$.

Using this specification, it is possible to integrate out $\boldsymbol{\beta}$ and σ^2 analytically by setting them equal to their analytic posterior mean, as a function of the remaining hyperparameters. Numerical optimisation determines the posterior mode of the remaining hyperparameters $\boldsymbol{\delta}$ and σ_ε^2 by optimising the log posterior density. The estimates for $\boldsymbol{\beta}$ and σ^2 are then calculated as a function of the MAP estimates for $\boldsymbol{\delta}$ and σ_ε^2 using the function for their analytic posterior means. This hyperparameter estimation is computationally intensive as evaluating the log posterior density requires the inversion of the GP covariance matrix with a N^3 computational cost. The optimisation for $\boldsymbol{\delta}$ and σ_ε^2 requires the computation of the log posterior density for each step. In addition to this, numerical optimisers can struggle to find the mode as the posterior surface has $P+1$ dimensions. This results in incorrect hyperparameter values and therefore incorrect estimates for the EVPPI. Most commonly, the optimiser gets stuck in a local optimum that entirely discounts the information from some components whilst inflating the δ^p

parameter for others. The can be mitigated by repeatedly finding the hyperparameters from different initial values but this increases the computation time still further.

In the next section, therefore, we present a fast Bayesian approximation method known as the Integrated Nested Laplace Approximation (INLA) that can be adapted to allow us to calculate the hyperparameters for a GP efficiently.

2.7 THE INTEGRATED NESTED LAPLACE APPROXIMATION

The Integrated Nested Laplace Approximation (INLA) is a fast approximate Bayesian inference method for a wide class of models known as latent Gaussian models [102]. This method performs significantly better than MCMC methods for a fixed computational cost and can therefore be used to quickly estimate the posterior distributions for parameters that characterise models of this class. Although this method could been seen as restrictive, latent Gaussian models can be used in many standard modelling scenarios, including regression models, dynamic models and spatio-temporal models [102].

A model can be expressed as a latent Gaussian model, if the data, y_i are defined by a parametric family with a parameter μ_i that is linked, through some function $h(\cdot)$, to a structured linear predictor η_i,

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f_j(\gamma_{ji}) + \sum_{k=1}^{n_\chi} \chi_k \gamma_{ki} + \epsilon_i, \tag{15}$$

where $f_j(.)$ are unknown functions of the model parameters, χ_k are fixed regression coefficients, ϵ_i is some error term, and n_f and n_χ are the number of functions of covariates and regressed covariates in the model [80]. The functions f_j can be of any form and typically can represent autoregressive models, spatial effects or seasonal effects. In addition to this flexibility, also note that not including the functions f_j retrieves a standard generalised linear model.

After specifying the structured linear predictor and parametric model, the parameters of this model must follow a specific structure that allows for fast computation of the posterior distributions for these parameters. It must be possible to write our model using a hierarchical structure with 2 layers of hyperparameters. At the first level of the hierarchy, we have a small number (typically no more than 6) of hyperparameters;

$$\mathbf{\Lambda} \sim \pi(\mathbf{\Lambda}), \tag{16}$$

that can theoretically follow any prior distribution with known parameters. The parameters that control the linear predictor, $\mathbf{\Gamma} = (\boldsymbol{\chi}, \boldsymbol{\eta}, \boldsymbol{\alpha})$, must then be a Gaussian Markov Random Field (GMRF) [100] dependent on these hyperparameters, $\mathbf{\Lambda}$,

$$\mathbf{\Gamma}|\mathbf{\Lambda} \sim N(\boldsymbol{\mu}(\mathbf{\Lambda}), \boldsymbol{Q}^{-1}(\mathbf{\Lambda})). \tag{17}$$

This implies that the parameter set $\boldsymbol{\Gamma}$ follows a normal distribution with a mean function and covariance function dependent on the hyperparameters. The precision matrix, $\boldsymbol{Q}(\boldsymbol{\Lambda})$, must be sparse where this sparsity corresponds to the *Markov* property of the field. An entry in this matrix is 0 if and only if the two entries are independent conditional on all other entries. This can be thought of a *sparse* neighbourhood structure underlying the parameters, where only parameters that are connected have a non-zero entry in the precision matrix. Thus, the precision matrix is sparse, allowing for fast computation of the posterior distribution for the parameters. Note that a sparse precision matrix implies *conditional* independence, whereas a sparse covariance matrix implies marginal independence, a much stronger restriction.

2.7.1 *The INLA algorithm*

INLA approximates the posterior distribution of the hyperparameters $\boldsymbol{\Lambda}$ by a nested Laplace approximation [102]. Operationally, INLA explores the approximate joint posterior of the hyperparameters by determining the density using the Laplace approximation at different points in the support of $\boldsymbol{\Lambda}$. Linear interpolation is then used to find the approximate posterior joint density for $\boldsymbol{\Lambda}$. The different points are found by "stepping" through the space by some fixed interval along the axis until the density falls below some known threshold. This creates a lattice of points that are then used to approximate the posterior marginals of $\boldsymbol{\Lambda}$ by numerical integration.

The marginals for the parameters $\boldsymbol{\Gamma}$ are then approximated using another (simplified) Laplace approximation. This Laplace approximation is evaluated at the hyperparameter values at each point in the lattice and the marginals for $\boldsymbol{\Gamma}$ are a weighted sum of the Laplace approximation for each configuration of the hyperparameter set (weighted by the density at that point). In this sense, the approximate marginals for $\boldsymbol{\Gamma}$ are *nested* within the Laplace approximation for posterior distribution of the hyperparameters.

INLA is fast and accurate, when compared to a fixed length MCMC sampling algorithm, provided the model structure is relevant. It is important to stress that INLA does not sample from the posterior distributions but numerically approximates the required distributions. This approximation can be improved by "stepping" through the hyperparameter space in finer steps. This will increase the computational time as more lattice points are used for both the numerical integration and to determine the distribution of the marginals for $\boldsymbol{\Gamma}$.

`R-INLA` [103] is an `R` package that allows for simple implementation of the INLA algorithm. It provides a simple user interface for this fast approximate Bayesian Inference and all results presented using the INLA algorithm are taken using this interface. Additional information about the INLA algorithm and the `R-INLA` interface can be found in [78–80, 101, 102].

2.8 SPATIAL STATISTICS AND STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS

An interesting application of GPs is in spatial statistics, where measurements are taken at different points in a spatial domain and distance has a physical interpretation. For example,

cases of influenza at locations in a geographical area (e.g. a country) or the level of pollution at different monitoring sites. In this setting, the main assumption is that points that are "closer" to each other in a geographical sense share more common features and are influenced by common factors than those "further away" [116].

A very popular specification of a spatial model when exact locations are available is based on the Matérn family of covariance functions [36], defined by

$$\mathcal{C}_{\mathrm{M}}(r,s) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa\|\boldsymbol{\phi}^r - \boldsymbol{\phi}^s\|)^{\nu}\mathrm{K}_{\nu}(\kappa\|\boldsymbol{\phi}^r - \boldsymbol{\phi}^s\|),$$

where $\boldsymbol{\xi} = (\sigma, \kappa, \nu)$ is a vector of hyperparameters, $\|.\|$ denotes the Euclidean distance and K_{ν} is the modified Bessel function of the second kind and order ν. The Matérn covariance function is related to the covariance function in (13), which can be obtained when δ^p is constant for all $p = 1, \ldots, P$ and $\nu \to \infty$ [98]. This implies that the resulting covariance matrix for a specific set of input values is still dense and that the responses are not conditionally independent.

Lindgren et al. [74] demonstrate that a sparse matrix can be used to approximate a GP with a Matérn covariance function (Matérn GP) by using Stochastic Partial Differential Equations (SPDE). It can be shown that whilst a Matérn GP is defined in terms of a relationship with the multivariate Gaussian it is also exactly equal to the function $g_t(\boldsymbol{\phi})$ that solves a specific SPDE

$$\tau(\kappa^2 - \Delta)^{\frac{\alpha}{2}} g_t(\boldsymbol{\phi}) = \mathcal{W}(\boldsymbol{\phi}),$$

where $\mathcal{W}$ is Gaussian noise, Δ is the Laplacian operator, $\alpha = \nu + \frac{P}{2}$ (with $P = 2$, in the spatial context) and the marginal variance of the Matérn GP is

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{\frac{P}{2}}}\kappa^{-2\nu}\tau^{-2}.$$

Therefore, finding the solution to this SPDE is exactly equivalent to finding the function $g_t(\boldsymbol{\phi})$, which as mentioned in §2.6 is instrumental in estimating the EVPPI.

The fundamental implication of this result is that efficient algorithms for solving SPDEs can be used to approximate the Matérn GP. In practice, the SPDE is solved using the finite element method [22]. First, the region over which the SPDE is being solved, i.e. the range of $\boldsymbol{\phi}$, is split into small areas, e.g. in the 2-dimensional case a grid of small triangles. An example of this triangulation in a true spatial example relating to rainfall in Switzerland is shown in Figure 7. There are two triangulation boundaries with the border of Switzerland added for clarity in this setting. An inner boundary encases (or "hugs") all the data points relatively tightly while the outer boundary is further from the points. This is because boundary conditions are imposed on the SPDE solver and we do not wish these to impact the estimation of the Matérn GP. The value of the Matérn GP is approximated by simple (linear) functions within each small triangular area. Therefore, the points are closer together within the inner boundary to give a good approximation to the Matérn GP, while in the outer region the grid can be rougher as the approximation is not as important.

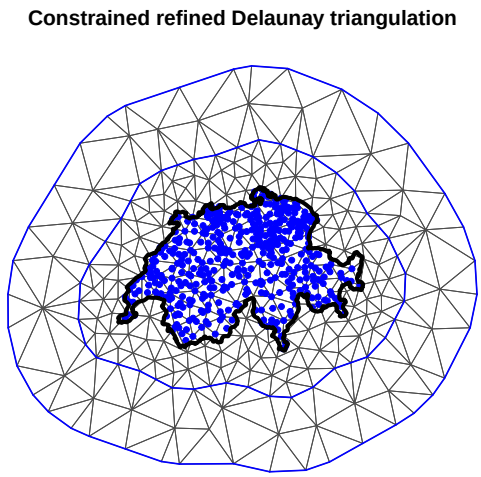


Figure 7: An example of the grid approximation used to approximate the Matérn GP in a spatial problem. The thick black line represents the border of Switzerland. The blue dots represent the positions where data points have been observed. These data points are used to estimate the value of the Matérn GP throughout the geographical space (i.e. the whole area covered by Switzerland, in this case).

In the 2-dimensional case, the approximation is given by the sum of weighted linear functions of $\boldsymbol{\phi}$ that are equal to 1 at one vertex (intersections where several triangle meet) and 0 at all other vertices. Therefore, the weights of the linear functions determine the value of the function at each vertex, where one of the linear functions is equal to 1 and all other linear functions are equal to 0. Between the vertices (and therefore within the triangles) the value of the Matérn GP is approximated by linear interpolation of the weights at the three corners of the triangle.

The weights entirely determine the value of the Matérn GP at all points in the region of interest and must be estimated from the data. Therefore, the weights are hyperparameters controlling the value of the function $g_t(\boldsymbol{\phi})$. The relationship between the weights is controlled by τ and κ. Crucially, the weights are assumed to have a multivariate Gaussian distribution with a specific precision matrix [74] which at least to a very good degree of approximation is sparse, since non-zero entries correspond loosely only with points that "share a triangle". This allows us to approximate a Matérn GP with a Latent Gaussian Field where the model in (14) becomes $\mathbf{NB}_t \sim \text{Normal}(\boldsymbol{H\beta}, \boldsymbol{\mathcal{C}}_{\text{M}} + \sigma_\varepsilon^2 \boldsymbol{I})$, which can be equivalently re-expressed as

$$\mathbf{NB}_t \sim \text{Normal}(\boldsymbol{H\beta} + f(\boldsymbol{\omega}), \sigma_\varepsilon^2 \boldsymbol{I}), \tag{18}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_G)' \sim \text{Normal}\left(\mathbf{0}, \boldsymbol{Q}^{-1}(\tau, \kappa)\right)$, where G is the number of grid points, is the contribution from the Matérn covariance matrix to the expected value of the net benefit and $\boldsymbol{Q}(\tau, \kappa)$ is the sparse precision matrix determined by the SPDE solution.

2.8.1 *Computing the EVPPI using SPDE-INLA*

This SPDE-INLA method has been developed and successfully applied in a spatial context [14, 15, 105], where inputs are proper coordinates (i.e. longitude and latitude, hence defined in a 2-dimensional space. However, calculating the EVPPI relies on a set of much higher dimensional inputs. While in theory the SPDE machinery works in higher dimensional spaces, the computational advantages will diminish in these cases.

To fully exploit the computational savings of the SPDE-INLA procedure, we re-express the problem of computing the EVPPI in a "spatial" context. In this case, the simulated parameter vector for $\boldsymbol{\phi}$ designates a point in the P-dimensional parameter space. We consider that the net benefit, calculated as a function of $\boldsymbol{\phi}$, has been "observed" at this point. We then wish to find a representation of these P-dimensional points in at most 2-dimensional space, so that we can efficiently estimate the Matérn covariance of this representation using the SPDE-INLA methodology. If no such projection exists, then the computational complexity of fitting a Matérn GP with a high dimensional SPDE means that the standard GP method is more appropriate.

As this projection will be used to predict the net benefit values, it makes sense to use a regression-based dimension reduction method. This class of methods tries to find a *sufficient* reduction, i.e. one for which the projection contains all the relevant information about the net benefit function. Formally, we can express this condition as $p\left(\mathbf{NB}_t \mid \boldsymbol{\phi}\right) = p\left(\mathbf{NB}_t \mid R(\boldsymbol{\phi})\right)$, where $R(\cdot)$ is the reduction function from $\mathbb{R}^P \rightarrow \mathbb{R}^d$ where P is the number of parameters in $\boldsymbol{\phi}$ and d is the minimum number of dimensions needed to capture all the relevant information in $\boldsymbol{\phi}$. Possible values for d can be determined by relatively standard testing procedures. There are a wealth of methods that can be used to estimate this sufficient reduction [29–32, 72, 73]. Specifically, we focus on Principal Fitted Components [30, 31] to calculate the EVPPI.

2.8.1.1 *Principal Fitted Components*

Principal Fitted Components (PFC) is a model based inverse regression method. This means that in order to find a sufficient reduction we consider a model for $\boldsymbol{\phi}$ as a function of $\mathbf{NB}_t$. As different models can be specified and lead to different sufficient reductions, the best fit model amongst a set of candidates should be chosen before finding the sufficient reduction.

The general form of these PFC models is based on a linear structure

$$\boldsymbol{\phi} = \boldsymbol{\mu} + \mathbf{Y}\boldsymbol{f}(\mathbf{NB}_t) + \boldsymbol{\epsilon},$$

where $\boldsymbol{\mu}$ is the intercept; $\mathbf{Y}$ is a $P \times d$ matrix to be estimated to determine the sufficient reduction; $\boldsymbol{f}(\cdot)$ is a vector-valued function of $\mathbf{NB}_t$; and $\boldsymbol{\epsilon}$ is an error term in the exponential family. The form of the error structure changes the way in which the reduction is calculated and methods have been developed for independent, heteroskedastic and unstructured errors.

In order to use PFC, the function $\boldsymbol{f}(\cdot)$ and the error structure need to be specified. Specifically, we consider normally distributed errors and set $\boldsymbol{f}\left(\mathbf{NB}_t\right) = \left[\alpha_1\mathbf{NB}_t, \alpha_2\mathbf{NB}_t^2, \ldots, \alpha_h\mathbf{NB}_t^h\right]^T$, although in general $\boldsymbol{f}(\cdot)$ can map to any function of $\mathbf{NB}_t$. Additionally, the number of dimen-

sions d needed to capture all the relevant information in $\boldsymbol{\phi}$ needs to be specified. It is then advisable to select the value of d associated with the best performing inverse regression specification, e.g. in terms of model fitting, as measured by information criteria such as the AIC [31].

The cost of fitting any of these models individually is negligible and thus fitting a number of models in correspondence to a set of chosen d values adds very little to the computational time required to estimate the EVPPI. In any case, because PFC assumes that the number of dimensions needed to capture the information in $\boldsymbol{\phi}$ is no larger than then number of dimensions in the function $\boldsymbol{f}(\cdot)$, for simple relationships between the net benefit and $\boldsymbol{\phi}$, the sufficient regression is low dimensional.

If one dimension is sufficient to capture all the information in $\boldsymbol{\phi}$, there is no harm in using a second component because this will only add information and the reduction will remain sufficient. On the other hand, using two dimensions when the AIC suggests $d > 2$, will lead to a loss of information. Consequently, the EVPPI estimate based on a two-dimensional reduction of $\boldsymbol{\phi}$ may be biased. In light of the large computational savings and the fact that the AIC has a tendency to overestimate d [31], it may still be worth using the projection to estimate the EVPPI and then perform thorough model checking (e.g. by means of the residual plots) to assess the performance of the GP model fit. We return to this point in §4.3.

From the theoretical point of view, PFC provides a robust method for determining the sufficient reduction [31]. Additionally, due to the flexibility of the INLA algorithm it is possible to cater for more complicated structures in the relationships between the net benefit and the parameters of interest. Both residual plots and extensions to the INLA algorithm are discussed when the practical implementation of this method presented in §4.3.

2.9 TESTING THE INLA-SPDE GAUSSIAN PROCESS REGRESSION METHOD

To assess this fast GP regression method, we present two case studies of health economic models and compare the estimates of the EVPPI obtained using the direct GP regression implemented by Strong *et al.* [108] and our SPDE-INLA projection method. For both case studies, random subsets of between 5 and 16 important parameters were considered to compare the performance of the GP procedures — notice that this represents the standard range of parameter subsets that would be used practically for EVPPI calculation using GP [40, 111]. For each subset, the EVPPI was calculated using both methods and for a willingness-to-pay threshold of $k = 20\,000$ monetary units, say £. The computational time and EVPPI estimate is then recorded for both methods to allow a direct comparison.

The first case study is the Vaccine example presented in §2.4.1.1. This example has 28 key parameters, as previously discussed, representing the probability of infection, the reduction in risk due to the vaccine, the occurrence of complications, the monetary costs of the interventions and the QALY loss due to different health states. However, there are also additional sources of uncertainty present in the model; for instance, the true number of people getting

influenza or the true number of people getting side effects. Considering all the uncertain quantities in the model, the number of parameters increases to 62.

The second case study is a simple fictional decision tree model with correlated parameters, presented at the `SAVI` web application [109] (hence this example is referred to as the "SAVI study"). The model has two treatment options and 19 underlying parameters. A more in-depth model description is presented in [12]. Most importantly, the 19 underlying parameters are assumed to follow a multivariate Gaussian distribution with known structure and thus the single loop MC methods can be used for some parameter subsets. The `SAVI` web application provides $10\,000$ PSA samples of all the 19 parameters, along with the simulated costs and benefits of both treatment options. This number of PSA samples poses a significant challenge for the standard GP regression method. Therefore, only $1\,000$ observations are used for the comparison with our SPDE-INLA method.

2.9.1 *Computational Time*

We begin our discussion of the two EVPPI estimation methods by comparing the computational time required to obtain an estimate. We used $1,000$ PSA samples for both case studies and both methods. To compare our SPDE-INLA method, we used the code available from Strong [108] with a slight modification. This modification changed the numerical optimiser for the hyperparameter estimation to give quicker computation time and more accurate results. In some cases this optimiser can struggle numerically and the slower optimiser must be used.

Additionally, to allow for a fair comparison between the two methods only 500 PSA runs were used to estimate the hyperparameters. It is suggested by Strong *et al.* [111] that the full PSA run should not used to calculate the hyperparameters to save computation time. Once the hyperparameters have been estimated, all $1,000$ PSA samples are used to find the fitted values $\hat{g}_t(\boldsymbol{\phi}_s)$, so all the information is utilised. Using all $1,000$ observations for the optimisation step can give more accurate results and is sometimes necessary (see for example §2.10).

The computational time for the GP regression increases substantially with the number of parameters of interest, between 17 and 470 seconds for the Vaccine case study and 17 and 71 seconds for the SAVI example. However, interestingly, the computation time does not increase uniformly for GP regression. This is due to the numerical optimisation as occasionally additional optimisation steps are required to reach convergence.

The computation time for our SPDE-INLA method remains constant as the number of parameters increases. This is because we are using projections and therefore the dimensionality of the GMRF is not increasing. The computation time of our SPDE-INLA method is significantly lower than the GP regression method, up to around 70 times faster. Even for 5 parameters of interest, it is between 2 and 2.5 times faster, despite the fact that we are using all the data points to estimate the hyperparameters, albeit from a projected input space.

To understand if our method scales to larger PSA datasets, EVPPI estimates using all $10\,000$ PSA samples from the SAVI case study were also calculated. The computational time required

Table 3: The computational time required (in seconds) to calculate an EVPPI using both the GP regression method and SPDE-INLA method for increasing numbers of parameters for both case studies

Number of parameters of interest	**Computation time (seconds)**			
	Vaccine Example		**SAVI Example**	
	GP	SPDE-INLA	GP	SPDE-INLA
5	17	9	17	7
6	42	10	14	7
7	45	10	18	7
8	57	11	21	8
9	74	8	26	8
10	86	8	31	9
11	70	7	37	8
12	60	8	47	8
13	84	11	52	7
14	188	8	66	6
15	470	7	70	7
16	121	8	71	7

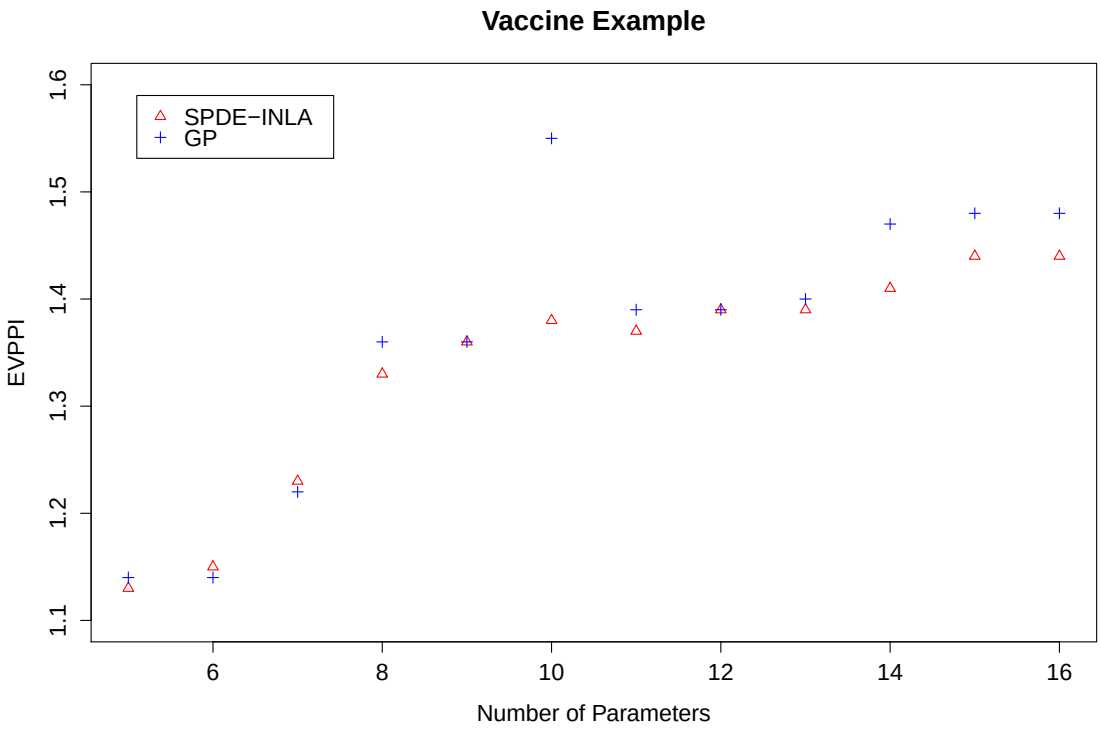
to calculate an EVPPI estimate was between 40 and 80 seconds with an average time of 56 seconds. This is important as it demonstrates that the computation time does not increase exponentially using the SPDE-INLA method; the computation time is between than 6 and 10 times slower for a 10 fold increase in number of PSA samples. Crucially, the speed of our SPDE-INLA method depends on the density of the SPDE grid approximation. Therefore, its computational effort could be decreased by using a sparser grid, although this would affect the quality of the EVPPI estimate. It would, therefore, be possible to use our method to calculate the EVPPI for larger PSA data sets. This may be relevant, for instance, in models involving individual level simulations [119], where larger PSA samples are required to fully assess the underlying distributions.

2.10 ASSESSING THE ACCURACY OF THE EVPPI ESTIMATORS

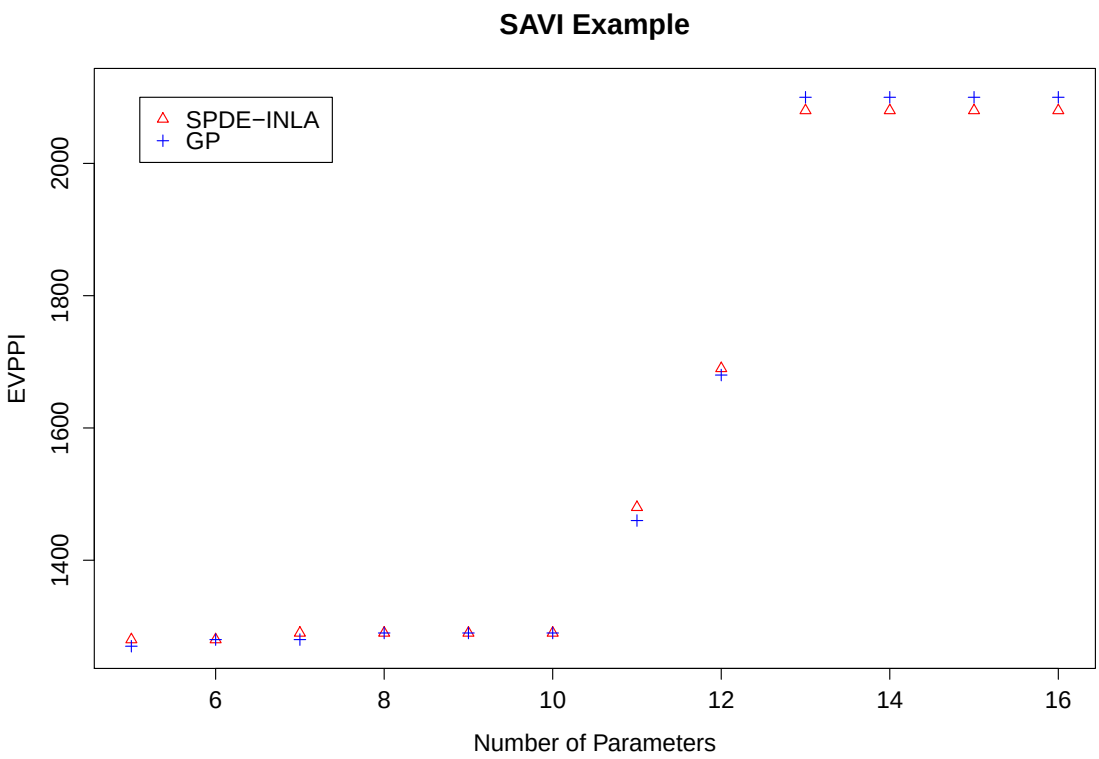
In general, it is difficult to establish whether an estimate of the EVPPI is accurate because analytical results are rarely available and MC simulation is very expensive. Thus, it is difficult to determine which method is more accurate when the two approximate EVPPI values diverge, as no baseline comparator is easily available. Nevertheless, there are at least two potential features that we can use to assess the reliability of our estimates.

2.10.1 *Monotonicity with respect to the number of parameters of interest*

We have seen that the EVPPI is a non-decreasing function of the number of parameters of interest, see §1.1. This means that, provided the smaller subsets are entirely contained within



(a)



(b)

Figure 8: The EVPPI estimate for the Gaussian Process regression method (GP) and the new method developed in this paper (SPDE) for increasing parameter subset size for the Vaccine (panel a) and the SAVI (panel b) case studies.

the larger subsets, the EVPPI estimates should be non-decreasing. This property provides one way to assess the accuracy of the methods: if one method fulfils this property and the other does not, then the former is likely to be more accurate. It is important to note that monotonicity is a necessary condition for a good EVPPI estimate but not sufficient. It is possibly to construct a function that gives a monotone sequence, such as simply giving the number of parameters in the set, but clearly does not estimate the EVPPI at all.

Figure 8 shows the EVPPI estimate for increasing parameter subset sizes for both case studies. For the Vaccine example, shown in panel (a), the standard GP regression method has some difficulty retrieving monotonicity, specifically for the parameter set containing 10 parameters which is clearly overestimated. Our SPDE-INLA method also overestimates the EVPPI for the set containing 10 parameters, but by 0.01 or less than 1%. This gives an indication of the accuracy of our EVPPI estimation method. This over-estimation for the standard GP method is due to incorrect estimation of the hyperparameters based on the reduced PSA dataset. If the hyperparameters are obtained using all $1\,000$ PSA samples then the EVPPI is 1.39, which respects monotonicity but the computation time increases to 256 seconds compared to 86.

For the SAVI example (panel b) the monotonicity is respected across both methods and the EVPPI values, rounded to 3 significant figures, are similar. For both examples, the EVPPI values for the smaller parameter subsets are similar across both methods. As the length of $\boldsymbol{\phi}$ increases the SPDE-INLA method underestimates slightly compared to the standard GP but only by at most 3% of the total EVPPI value. There is evidence that the SPDE-INLA method is accurate while possibly guarding against spurious results that come from estimating the hyperparameters based on a smaller subset of the PSA samples. Although, the SPDE-INLA method may slightly underestimate the EVPPI for larger subsets.

2.10.2 *The SPDE-INLA method as an EVPI approximation*

Therefore, to investigate whether our method underestimates the EVPPI for these larger parameter subsets, we use the method to calculate the overall EVPI, which represents the largest parameter subset available for each example. As mentioned in §1.4.1, the overall EVPI (4) can easily be estimated by MC simulation. We can then use our method to calculate the overall EVPI, by considering that all the underlying model parameters are of interest. This allows us to compare our method directly with the "true" MC EVPI estimate. The EVPI was calculated for both cases studies and the results are shown in Table 4. The computational time required to calculate these estimates is 14 seconds for the Vaccine example and 8 seconds for the SAVI.

For both case studies, the SPDE-INLA approximation is correct to two significant figures, with a small discrepancy in the third significant figure. This gives a further indication that the EVPPI estimated using our method is an accurate method for calculating the EVPPI, possibly demonstrating that even for large numbers of parameters of interest the underestimation of the EVPPI is not severe.

Table 4: The EVPI values calculated using the PSA samples directly and our SPDE-INLA method

Case Study	MC EVPI	SPDE-INLA EVPI
Vaccine	2.52	2.51
SAVI	2100	2080

For the Vaccine example, there are 62 parameters that contribute to the model uncertainty. We are therefore approximating the Matérn covariance function with a projection from a 62- to a 2-dimensional space. However, despite the difficulty of preserving the original data structure with this projection, the AIC suggests that this is a sufficient reduction and the EVPI estimate is still very close to the true value.

2.10.3 *Analytic Results*

Finally, we can compare our method with the "true" MC values calculated for the SAVI example. Strong et al. [111] provide these EVPPI values for three different parameters subsets of size 2 and 4. Both GP regression and GAM can be used to calculate the EVPPI using all $10\,000$ available PSA samples and the willingness-to-pay is fixed at £$10\,000$. Table 5 gives the "'truth" together with all three estimated values, GAM, GP regression and SPDE-INLA. It is clear that our method's performance is in line with the other approximation methods while all 3 methods overestimate the true EVPPI value in this setting.

Table 5: Comparison of the EVPPI estimation methods, standard GP, GAM regression and the SPDE-INLA method with "true" EVPPI values based on 10^7 Monte Carlo simulations.

	EVPPI estimate (Time to compute in seconds)			
Parameter Subset	**MC Simulations**	**GP**	**GAM**	**SPDE-INLA**
2 Parameters - ϕ_5, ϕ_{14}	248	$274(375)$	$277(0.78)$	$278(41)$
4 Parameters - $\phi_5, \phi_6, \phi_{14}, \phi_{15}$	841	$861(367)$	$862(98)$	$856(48)$
2 Parameters - ϕ_7, ϕ_{16}	536	$549(390)$	$546(0.25)$	$549(43)$

The computational time required to calculate these estimates are given in Table 5. Clearly, for the 2 parameter setting the GAM regression method is the most appropriate as it takes under a second to calculate the EVPPI estimate. However, for the 4 parameter example, the computational time is lowest for our SPDE-INLA method, which takes half the computational effort of the GAM regression method and around 10% of the standard GP. This also demonstrates that for larger PSA samples, GAM regression may be preferred over GP regression in terms of computational time, even for larger subsets.

2.10.4 *Standard Error*

Thus far, we have not compared the standard error of the SPDE-INLA method with the standard error of the alternatives. It would be possible to calculate the standard error for the SPDE-INLA method using a similar procedure as that outlined for the original formulation of the GP regression. Specifically, the SPDE-INLA model could be used to simulated predicted values from a multivariate Gaussian distribution. These simulated values could then be used to estimate the EVPPI a large number of times and the standard deviation of these estimates would give an approximation to the standard error of the EVPPI estimate.

However, as in the standard GP case, this standard error would ignore uncertainty arising from the misspecification of the SPDE-INLA model. This may have more serious implications in the SPDE-INLA model as we are projecting from $\boldsymbol{\phi}$ to estimate the Matérn covariance and therefore ignoing uncertainty due to this projection may give an inaccurate representation of the uncertainty in the EVPPI estimate from the SPDE-INLA method.

2.11 CONCLUSION

This chapter began with a discussion of the available methods for calculating the EVPPI. A full analysis of the more recent methods was undertaken and it was concluded that the best EVPPI estimation methods are based on non-parametric regression [111]. While these methods are very efficient when only a small number of parameters of interest are considered, they become expensive in higher dimensional settings. This is especially true when coupled with large numbers of PSA simulations. Therefore, we developed an approximation method for fast Gaussian Process regression to reduce the computation effort required to calculate the EVPPI.

This novel method is based on a spatial interpretation of GP regression and projections into 2-dimensional space. This, in turn, allows us to use a fast computation method developed in spatial statistics, based on finding a latent Gaussian field that approximates a Matérn GP. Finally, this allows us to use the INLA methodology for fast Bayesian computation to fit the GP. It also allows us to find fitted values at no additional cost, which are then used to estimate the EVPPI.

There seems to be little loss of accuracy when using our method. For larger subsets of parameters of interest, the EVPPI estimate is slightly underestimated compared to the standard GP methods, but this does not seem to be severe as the EVPI estimates are very close. Additionally, in some examples, our method seemed to be more accurate than the standard GP regression method, as we avoid breakdowns in numerical optimisation. These results are conditional on the dimension reduction being sufficient to capture all the information in $\boldsymbol{\phi}$.

3

ESTIMATION METHODS FOR THE EXPECTED VALUE OF SAMPLE INFORMATION

This chapter begins with a brief overview of the currently available EVSI estimation methods to determine the required elements for the novel EVSI estimation method. We then discuss an EVSI estimation method based on moment matching and elements of preposterior analysis. We present this method using several examples of varying complexity, each exploring a different aspect of the method. The chapter then concludes with an extension to this methodology using Bayesian non-linear regression to estimate the EVSI for different sample sizes. This allows us to present a final example which uses the Expected Net Benefit of Sampling (ENBS) to determine optimal study design in a practical health economic evaluation.

The methods in this chapter have been written up in three papers. Firstly, the main method is published in *Medical Decision Making* in a paper entitled *Efficient Monte Carlo Estimation of the Expected Value of Sample Information using Moment Matching* [62]. The technicalities regarding the specific simulation size required to accurately estimate the EVSI are from the paper entitled *Calculating the Expected Value of Sample Information using Efficient Nested Monte Carlo: A Tutorial*, submitted to *Value in Health* [57]. Finally, the Bayesian non-linear regression methodology is intended for submission in *Medical Decision Making* in a paper entitled *Estimating the Expected Value of Sample Information Across Different Sample Sizes using Moment Matching and Non-Linear Regression* [61].

3.1 CURRENT METHODS FOR EVSI ESTIMATION

To discuss the currently available approximation methods for the EVSI, first recall its definition:

$$\text{EVSI} = \mathsf{E}_{\boldsymbol{X}}\left[\max_t \mathsf{E}_{(\boldsymbol{\theta}|\boldsymbol{X})}\left[\text{NB}_t(\boldsymbol{\theta})\right]\right] - \max_t \mathsf{E}\left[\text{NB}_t(\boldsymbol{\theta})\right]. \tag{19}$$

As with the EVPPI, the MC estimator of the EVSI requires nested simulations and it is the inner "posterior" expectation that makes this estimator computationally intensive. Therefore, EVSI estimation methods have been developed to approximate this inner expectation without using full MC simulations.

As the EVSI is concerned with estimating the value of a specific study design, the estimation methods for the EVSI are much more varied than for the EVPPI. Specifically, some of the methods are concerned with estimating the EVSI for a specific type of data collection exer-

cise, such as cluster randomised clinical trials [1, 11, 21, 120]. In general, these methods are likely to be accurate and computationally efficient in the settings for which they were designed. However, they have limited applicability and are typically based on finding analytical expressions/approximations for the posterior mean of the net benefit under specific assumptions about the parameters and the sampling distribution.

For example, if the PSA distribution for the parameters $\boldsymbol{\theta}$ and the sampling distribution of the data are from a conjugate family, it is possible to find the "posterior" for the parameters analytically, rather than resorting to simulation, and therefore it may be possible to find an analytical expression for the mean of the posterior net benefit. In these settings, the EVSI can be calculated using a one-step procedure, where each PSA simulation $\boldsymbol{\theta}_s$ for $s = 1, \ldots, S$ is used to simulate a single potential sample $\boldsymbol{X}_s$ from $p(\boldsymbol{X} \mid \boldsymbol{\theta}_s)$. These ideas can be extended when the design of the experiment and the net benefit functions conform to certain conditions. In addition to this, some researchers use the central limit theorem to justify an approach based on exploiting normal-normal conjugacy [43, 122].

In general, however, we are more interested in general purpose estimation methods for the EVSI. These are methods that can be used irrespective of the underlying structure of the health economic model and the design of the future experiment. These methods are all based on approximating the inner conditional expectation in equation (6).

Firstly, Sweeting and Kharroubi [114] developed a method to approximate the posterior expectation using a weighted sum of the posterior density evaluated at a small number of points. As the EVSI is based on the posterior expectation, this method can be used to approximate the EVSI [10, 70] using a small number of posterior evaluations. However, the formulæ are relatively complicated and assume that the posterior distribution can be maximised and differentiated, which limits applicability.

It is also possible to extend methods based on the *unit loss integral* (see §2.1) to calculate the EVSI [68]. This method begins by fitting a linear regression model between some of the model parameters and the incremental net benefit. Provided the parameters of interest can be approximately modelled using a normal distribution, it is then possible to estimate the mean and variance of the "posterior" expectation. The unit loss integral then calculates the EVSI as a function of this mean and variance. A similar approach has been developed in [67] using regression to take advantage of closed form expressions for the posterior mean and variance in normal-normal conjugate models, without relying on the unit loss integral. This extension does not require a normal distribution for either the INB or the model parameters.

More recently, Menzies [84] has created an EVSI calculation method based on a sampling importance re-sampling approach [99]. In this method, the PSA simulations for the net benefit are re-weighted based on the likelihood of the data for each of the simulations of the parameters $\boldsymbol{\theta}_s$. This method avoids rerunning the health economic model as it is based solely on the PSA simulations but it does involve a large number of matrix calculations which can be costly if the PSA simulation size is large. It also requires that the full sampling distribution of the data can be specified in order to re-weight the samples.

Finally, Strong *et al.* [112] developed a non-parametric regression method that is similar to the EVPPI calculation method presented in §2.3. For the EVSI, it is assumed that the data $\boldsymbol{X}$ can be summarised in a low-dimensional sufficient statistic. A sufficient statistic is then calculated for each $\boldsymbol{\theta}_s$ for $s = 1, \ldots, S$ and non-parametric regression is used to estimate the "posterior" mean based on each statistic. This method is fast but it can be challenging to determine a low-dimensional sufficient statistic for some studies.

In light of these methods, we are interested in developing an approach that can be used irrespective of the underlying health economic model. We also want to avoid limitations on the form of the data. Therefore, our method can be used in any setting where a Bayesian model can be specified for the parameters and the potential data. This also allows us to consider complex designs, something that has been lacking from a large number of previous methods [67], particularly, it is challenging to calculate the EVSI for unbalanced designs, trials targeting multiple different model parameters and trials with missingness or dropout.

Therefore, in this section, we develop an alternative approximation method for the EVSI by determining the properties of the *distribution of the preposterior mean* §3.2.1. We then demonstrate that it is possible to approximate this distribution using moment matching. To achieve this, we estimate the mean and variance of the preposterior mean using a small number of future posterior samples. We discuss the theoretical grounding for this estimation method to highlight scenarios where moment matching is unsuccessful at approximating the EVSI.

We compare our methodology with analytic results, where they are available, simulation based approaches and the EVSI estimation method based on sufficient statistics [112]. Specifically, we demonstrate that in the considered examples §3.5.3, our moment matching approach is successful in most settings where the sample size of the future data collection exercise is sufficiently large.

3.2 PREPOSTERIOR ANALYSIS

Preposterior analysis encompasses a large suite of approaches that are concerned with estimating the properties of a posterior distribution *before* relevant data have been collected. These approaches are used in many different domains, from model calibration [3, 69], to model checking [8] and experimental design [19, 44, 65, 118]. The EVSI is an important example of preposterior analysis as we are interested in the posterior expectation of the known-distribution net benefit *before* the data have been collected.

Therefore, to discuss this EVSI estimation method, we begin by formally framing the EVSI within a Bayesian context. Firstly, the PSA distribution $p(\boldsymbol{\theta})$ is the *prior* for the model parameters. Interestingly, the PSA distribution is normally informed by past data, e.g. in the form of a previously conducted trial, and would normally be denoted as $p(\boldsymbol{\theta} \mid \mathcal{D})$, where $\mathcal{D}$ indicates the existing data. But, as we are discussing the collection of future data, the information in $\mathcal{D}$ is *prior* to the information in $\boldsymbol{X}$. Note, therefore, that the *prior* contains a relatively large amount of information about the underlying model parameters and the known-distribution net benefits.

Throughout this section, we will also need to consider the *prior* for the known-distribution net benefits. To ease the explanation of the EVSI calculation method, we will use the notation $\mathsf{NB}_t(\boldsymbol{\theta}) = \mathsf{NB}_t^{\boldsymbol{\theta}}$, so the prior for $\mathsf{NB}_t^{\boldsymbol{\theta}}$ is denoted $p\left(\mathsf{NB}_t^{\boldsymbol{\theta}}\right)$.

As before, data that would arise from the study are denoted $\boldsymbol{X}$ and are defined using the sampling distribution $p(\boldsymbol{X} \mid \boldsymbol{\theta})$. The distribution of $\boldsymbol{X}$ is given by the *prior* predictive distribution – again noting that *prior* means "prior to the new sample" and is conditional on $\mathcal{D}$. In most settings, the data collection exercise will produce data about N different individuals. From this point forward, therefore, N will be referred to as the *sample* size and is reserved for discussion about $\boldsymbol{X}$. This contrasts directly with *simulation* size S which is reserved for discussion about simulating from the prior or posterior for $\boldsymbol{\theta}$. This distinction allows us to discuss S simulations of sample size N which relates to S simulations from the prior predictive distribution $p(\boldsymbol{X})$ of a data collection exercise with sample size N. Finally, $\boldsymbol{X}_s$ will be used to denote the s-th *simulated* data set, whereas $\boldsymbol{X}_N$ denotes a theoretical data collection exercise of *sample* size N.

3.2.1 *The distribution of the preposterior mean*

The *distribution of the preposterior mean* is the distribution over the possible future values for the posterior mean *before* the data have been collected. Therefore, the preposterior mean is exactly the inner posterior expectation in the definition of the EVSI (6). To understand this quantity and its distribution more clearly consider a standard Bayesian analysis, performed after the data have become available and observed to the value $\boldsymbol{x}$. In this case, the posterior mean will simply be a number (or vector for multivariate distributions), with the value of the mean conditional on those data $\boldsymbol{x}$. Therefore, if the data are unobserved and we consider $p(\boldsymbol{X})$ as the distribution over all possible datasets, then the posterior mean has a distribution dependent on $p(\boldsymbol{X})$.

As in standard Bayesian analysis, the preposterior mean of the net benefit is given by

$$\mu_t^{\boldsymbol{X}} := \mathsf{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right] = \int_{\boldsymbol{\Theta}} \mathsf{NB}_t^{\boldsymbol{\theta}}\, p(\boldsymbol{\theta} \mid \boldsymbol{X}) d\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} \mathsf{NB}_t^{\boldsymbol{\theta}}\, \frac{p(\boldsymbol{X} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\boldsymbol{X})} d\boldsymbol{\theta}.$$

The only difference for preposterior analysis is that $\boldsymbol{X}$ is a random variable rather than an observed data set $\boldsymbol{x}$. Throughout this chapter, we use the notation $\mu_t^{\boldsymbol{X}}$ to denote the preposterior mean, for decision $t = 0, \ldots, T$ highlighting that the preposterior mean is a function of $\boldsymbol{X}$.

Based on this notation, the definition of the EVSI can be rewritten as

$$\mathsf{EVSI} = \mathsf{E}_{\boldsymbol{X}}\left[\max_t \mu_t^{\boldsymbol{X}}\right] - \max_t \mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right].$$

Notice that this definition requires that we determine the treatment with the largest posterior mean net benefit for each potential sample. Therefore, to calculate EVSI, it is necessary to find the distribution of the joint preposterior mean across all the alternative treatment options.

3.2.2 *Examples of the distribution of the preposterior mean*

To illustrate the concept of the preposterior mean, we introduce a simple example using the Beta-Binomial conjugate family. Assume that a new drug is available and is associated with a probability θ of curing a particular disease. As this drug is new, it is assumed that there is very limited evidence on its effectiveness. This assumption could be expressed, in a very simplistic way, by modelling $\theta \sim Beta(1,1)$. We note, however, that it is likely that some information would be available, in practical settings (e.g. from small trials), so a more informative prior should be used instead. The effectiveness measure is whether the disease has been cured, meaning that the population level effectiveness is θ. Assume further that the drug costs c are known and that the willingness-to-pay is some constant k.

The reference treatment option is to do nothing. This has no cost and no effectiveness as this (non-life-threatening) disease does not improve without drug intervention. This implies that the two net benefit values are

$$\mathsf{NB}_0^\theta = 0 \quad \text{and} \quad \mathsf{NB}_1^\theta = k\theta - c.$$

The future experiment is to give N people the drug and observe how many are cured. This can be expressed using a binomial distribution for the future sample $X \mid \theta \sim Bin(N,\theta)$. The prior predictive distribution in this setting is given by

$$\begin{aligned} p(X) &= \int_0^1 \binom{N}{X} \phi^X (1-\phi)^{N-X} d\phi = \mathcal{B}(X+1, N-X+1) \frac{N!}{X!(N-X)!} \\ &= \frac{X!(N-X)!N!}{(N+1)!X!(N-X)!} = \frac{1}{N+1}, \end{aligned}$$

where $\mathcal{B}(\cdot,\cdot)$ is the Beta function. This implies that all samples for X are equally likely. Note that this prior predictive distribution for X is entirely due to our choice of prior.

Once the prior predictive distribution for X is known, the distribution of the preposterior mean for the two treatment options is determined by calculating the posterior mean (conditional on X) for both the net benefit functions. Obviously, the posterior mean for NB_0^θ is $\mu_0^X = 0$ and therefore the distribution of the preposterior mean is simply a point mass at 0. However, the posterior mean for NB_1^θ does depend on the future value of the random variable X:

$$\begin{aligned} \mu_1^X = \mathsf{E}_{\theta|X}\left[\mathsf{NB}_1^\theta\right] &= \int_0^1 (k\theta - c)\ p(\theta \mid X) d\theta \\ &= k\frac{1+X}{2+N} - c, \end{aligned}$$

as $\theta \mid X \sim Beta(1+X, 1+N)$. Therefore, the distribution of μ_1^X is conditional on the uniform prior predictive distribution for X, which in turn is conditional on the uniform prior for θ. This means that the distribution of the preposterior mean is uniform over all possible values for μ_1^X, calculated as a function of the $N+1$ possible X values.

The EVSI can then be used to summarise μ_0^X and μ_1^X and, dependent on the values of k, c and N, gives an upper limit to the value for the future data collection exercise. We note that this trial has value as the future sample may change our optimal treatment. A future sample indicates that $t = 1$ (the new drug) is optimal if

$$\frac{1+X}{2+N} > \frac{c}{k}$$

and that $t = 0$ (doing nothing) is optimal otherwise. As an example, for $k = 20\,000$, $c = 10\,000$ and $N = 5$, the EVSI can be calculated exactly as:

$$\text{EVSI} = \left(0+0+0+\frac{10\,000}{7}+\frac{30\,000}{7}+\frac{50\,000}{7}\right)\frac{1}{6} - 0 = \frac{15\,000}{7} = 2\,142,$$

which must be compared with the cost of a trial with 5 participants to determine whether that specific trial is worth funding.

3.2.2.1 *Distribution of preposterior mean for Exponential-Gamma conjugacy*

In this second example, a Gamma prior is assumed for the model parameter $\theta \sim Gamma(\alpha, \beta)$. The data collection exercise is then assumed to be N independent observations from an Exponential distribution conditional on θ; $X_j \sim Exp(\theta)$ with $j = 1, \ldots, N$. We consider the distribution of the preposterior mean for different values of N, where the two net benefit functions are:

$$\text{NB}_0^\theta = c_0 \quad \text{and} \quad \text{NB}_1^\theta = k\theta - c_1.$$

Figure 9 presents the distribution of the preposterior mean for incremental net benefit ($\text{NB}_1^{\boldsymbol{\theta}} - \text{NB}_0^{\boldsymbol{\theta}}$) for $\alpha = 5$, $\beta = 1$, $k = 200$, $c_0 = 900$ and $c_1 = 100$.

Firstly, note that the distribution of the preposterior mean gets closer to the prior as the sample size increases for the data collection exercise. The distribution of the preposterior mean also has larger variance as the sample size increases. These two properties are at odds with the intuition that the distribution of the mean gets more concentrated and symmetrical as the sample size increases. Nonetheless, these counterintuitive results hold because the "strength" of the data increases as the sample size increases and so the posterior mean deviates further from the prior mean. This implies that the variance of the preposterior mean does increase as the sample size increases. In addition to this, at the current state of knowledge, i.e. *before* the data collection has taken place, it is not possible to learn anything additional about the parameters. Therefore, if a future data collection exercise gives exact information (perfect information) about the parameter location then the distribution of possible posterior means is equal to distribution of possible parameter values i.e. the prior distribution. To clarify these ideas further, we now consider a normal-normal conjugate example.

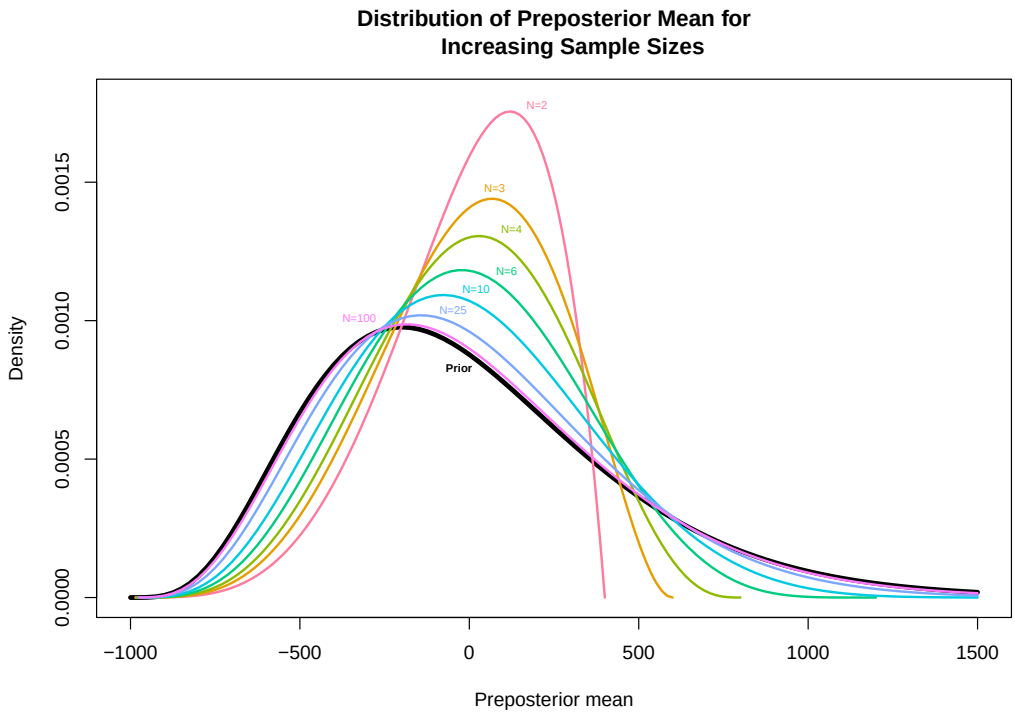


Figure 9: The distribution of the exact preposterior mean for different samples sizes using Exponential-Gamma conjugacy, with the prior for the net benefit marked in black.

3.2.2.2 *Distribution of the preposterior mean for Normal-Normal conjugacy*

For this example, a Gaussian prior for θ is defined as

$$\theta \sim N(\theta_0, \sigma_\theta^2)$$

while the data collection exercise is N independent samples from a different Gaussian distribution;

$$X_i \sim N(\theta, \tau^2).$$

To use conjugacy results, it is assumed that the variances σ_θ^2 and τ^2 are known. Therefore, the sample mean of $\boldsymbol{X}$ has a Gaussian distribution, conditional on θ and so the prior-predictive distribution for $\bar{\boldsymbol{X}}$ is

$$\bar{\boldsymbol{X}} \sim N\left(\theta_0, \sigma_\theta^2 + \frac{\tau^2}{N}\right).$$

Finally, the net benefit functions for this example are defined as

$$\mathrm{NB}_0^\theta = 0 \quad \text{and} \quad \mathrm{NB}_1^\theta = k\theta - c.$$

The preposterior mean for the $\mathrm{NB}_1^{\boldsymbol{\theta}}$ is then

$$\mathrm{E}_{\theta|\boldsymbol{X}}(\mathrm{NB}_1^{\boldsymbol{\theta}}) = k\left(\frac{\tau^2}{\tau^2 + N\sigma_\theta^2}\theta_0 + \frac{\sigma_\theta^2}{\frac{\tau^2}{N} + \sigma_\theta^2}\bar{\boldsymbol{X}}\right) - c,$$

which is a linear function of the prior predictive distribution for $\boldsymbol{X}$, which is Gaussian. Therefore, the distribution of the preposterior mean is Gaussian with mean and variance equal to the mean and variance of the preposterior mean;

$$\mathsf{E}_{\theta|\boldsymbol{X}}(\mathsf{NB}_1^{\boldsymbol{\theta}}) \sim N\left(k\theta_0 - c, k^2 \frac{\sigma_\theta^4}{\frac{\tau^2}{N} + \sigma_\theta^2}\right).$$

Thus, the distribution of the preposterior mean is centred on the prior mean for $\mathsf{NB}_1^{\boldsymbol{\theta}}$. So, it is not possible to learn any additional information about $\mathsf{NB}_1^{\boldsymbol{\theta}}$ on average before doing the study. Additionally, it is easy to see that as N increases, the denominator of the variance gets smaller. This implies that the variance increases as the sample size of the data collection exercise increases. Finally, as $N \to \infty$ the variance simplifies to the prior variance, demonstrating that as the sample size of $\boldsymbol{X}$ increases, the distribution of the preposterior mean $p(\mu_1^{\boldsymbol{X}})$ tends to the prior for the $\mathsf{NB}_1^{\boldsymbol{\theta}}$.

3.3 ESTIMATING THE DISTRIBUTION OF THE PREPOSTERIOR MEAN

In §1.4.3, we saw that the distribution of the preposterior mean can be calculated by MC simulation, also formalised outside of the health economic context [94]. We discussed that this method is computationally intensive as a large number of future samples $\boldsymbol{X}_s$ for $s = 1, \ldots, S$ should be simulated. Each sample is then used to update the prior and find the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{X}_s)$ by simulation in order to calculate the posterior mean. Our methodology presented below reduces the number of simulations required from $p(\boldsymbol{X})$ by exploiting the information available in the prior.

Throughout, we have been concerned solely with the *distribution* of the preposterior mean which, in §3.2.2, was computed by finding the prior predictive distribution $p(\boldsymbol{X})$ and the functional relationship between the preposterior mean and the sample $\boldsymbol{X}$. However, if the prior predictive distribution is not known, as in most practical situations, then a known functional form of the $\mu_t^{\boldsymbol{X}}$ cannot be used to determine the distribution of the preposterior mean. Therefore, while other EVSI estimation methods have focused on estimating a functional form for the expected net benefit *conditional* on the future sample (cf §3.1), we focus solely on estimating the *distribution* of the preposterior mean. Therefore, we are simply concerned with estimating a probability density, for which there is a large wealth of statistical theory.

3.3.1 *Expectation and Variance for the preposterior mean*

To approximate the probability density of the preposterior mean, we begin by estimating its mean and variance. In this analysis, the interest lies with the expectation and variance *conditional* on the value of $\boldsymbol{X}$, implying that standard formulæ for conditional iterated expectation can be used to calculate both the expectation and variance of the preposterior mean.

Therefore the mean of the distribution of the preposterior mean is given by

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

which implies that the expectation of the preposterior mean is equal to the prior mean, confirming what was observed in the examples. Again, preposterior analysis does not give any additional information about the net benefit. On average, over all the expected samples (which are conditional on the prior beliefs), the expected net benefit is the same.

The variance of the preposterior mean has a more complex formula but can also be re-expressed using iterated expectation as

$$\sigma_{\boldsymbol{X}}^2 = \mathsf{Var}_{\boldsymbol{X}}\left[\mu_t^{\boldsymbol{X}}\right] = \mathsf{Var}_{\boldsymbol{X}}\left[\mathsf{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right]\right] = \mathsf{Var}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right] - \mathsf{E}_{\boldsymbol{X}}\left[\mathsf{Var}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right]\right].$$

This means that the variance of the preposterior mean is equal to the variance of the prior known-distribution net benefit minus the expectation, over all the possible samples $\boldsymbol{X}$, of the posterior variance. Therefore, to calculate the variance of the preposterior mean distribution practically, the average posterior variance over all possible samples $\boldsymbol{X}$ must be estimated. However, §3.4.1 demonstrates that the average posterior variance can be estimated using a significantly reduced number of posterior samples compared to calculating the EVSI by simulation.

3.3.2 *Moment Matching*

Moment matching is a common method of performing parameter inference within a model but has recently been applied in the context of estimating an unknown density [18, 48]. In general, an unknown distribution can be accurately characterised by a large set of moments. However, it can also be approximated using a known distribution and “matching” a small number of the moments. This means that an alternative family of distributions is chosen and then parameters are found to determine a distribution in this family with the same moments as the distribution of interest. In the simplest setting, this involves approximating the distribution of the preposterior mean by a Gaussian with the mean and variance calculated using the formulæ in (20).

However, this is unlikely to be sufficiently accurate for our purposes as the EVSI is strongly influenced by the tails of the distribution of the preposterior mean. This is due to the fact that the optimal decision is most likely to be different from the current optimal decision in the tails as the current decision is, by definition, optimal for the majority of the prior mass.

Therefore, the EVSI estimate will be significantly improved if the distribution of the preposterior mean is approximated using moment matching with an alternative distribution that is closer to that of the preposterior mean. In fact, we suggest that the prior distribution for the net benefit is similar enough to the true distribution of the preposterior mean to give a good approximation for the EVSI, specifically for larger sample sizes N. This is because, while a specific future sample would give additional information, the preposterior analysis (before the data are collected) cannot give any information in addition to that contained in the prior. At

this point, it is worth reiterating that the *prior* is normally a posterior distribution conditional on data and is, therefore, more likely to contain useful information about the parameters and the net benefits.

In this sense, the unknown distribution of the preposterior mean net benefit $p(\mu_t^{\boldsymbol{X}})$ is approximated by a distribution with the correct expectation and variance but all other distributional properties such as the skewness are determined from the prior $p(\mathsf{NB}_t^{\boldsymbol{\theta}})$. This idea can be generalised to situations where the sample $\boldsymbol{X}$ is dependent on a subset $\boldsymbol{\phi}$ of the underlying model parameters $\boldsymbol{\theta}$ (see §3.4.3) and although we acknowledge difficulties in some specific cases (§3.5.1), it is successful in many settings (§3.5.3).

3.3.2.1 *Linear transformation to moment match*

Practically, to "moment match" with the prior, the distribution of the preposterior mean is estimated by a shifted and rescaled version of the prior. This implies that a linear transformation of $\mathsf{NB}_t^{\boldsymbol{\theta}}$ must be found such that $a\,\mathsf{NB}_t^{\boldsymbol{\theta}} + b$ has the same mean and variance as the distribution of the preposterior mean:

$$\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\left[a\,\mathsf{NB}_t^{\boldsymbol{\theta}} + b\right] &= \mathsf{E}_{\boldsymbol{\theta}}\left[\mu_t^{\boldsymbol{X}}\right] \Rightarrow a\mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right] + b = \mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right] \\
\mathsf{Var}_{\boldsymbol{\theta}}[a\,\mathsf{NB}_t^{\boldsymbol{\theta}} + b] &= \mathsf{Var}[\mu_t^{\boldsymbol{X}}] \Rightarrow a^2\mathsf{Var}_{\boldsymbol{\theta}}[\mathsf{NB}_t^{\boldsymbol{\theta}}] = \sigma_{\boldsymbol{X}}^2,
\end{aligned}$$

where $\sigma_{\boldsymbol{X}}^2$ is the variance of the preposterior mean distribution that can be written as a function of the prior variance and the expected posterior variance. Solving for a and b yields

$$a = \sqrt{\frac{\mathsf{Var}_{\boldsymbol{X}}\left[\mu_t^{\boldsymbol{X}}\right]}{\mathsf{Var}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right]}} = \frac{\sigma_{\boldsymbol{X}}}{\sqrt{\mathsf{Var}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right]}} \quad \text{and} \quad b = \mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right](1-a), \tag{20}$$

which depend on the prior expectation, prior variance and expected posterior variance for the net benefit.

Interestingly, these constants allow for a relatively simple interpretation of the approximation of the density of the preposterior mean. The constant a can be thought of as the proportion of the variance in $\mathsf{NB}_t^{\boldsymbol{\theta}}$ that is *explained* by the future sample $\boldsymbol{X}$. This means that the more closely the sample $\boldsymbol{X}$ reflects the underlying $\boldsymbol{\theta}$ values, the higher the value of a.

The constant b in (20), however, is the prior (and preposterior) mean weighted by one minus this *explained variance*. This weight is directly related to how closely the sample reflects the underlying values of $\boldsymbol{\theta}$. Thus, the density of the preposterior mean is estimated as a convex combination of the prior for the net benefit and the mean of the net benefit.

As the sample size in the data collection exercise increases, then the sample $\boldsymbol{X}$ (or some summary measure of $\boldsymbol{X}$) reflects the underlying value of $\boldsymbol{\theta}$ more closely. In turn, this implies that the density of the preposterior mean reflects the prior for the net benefit more and more closely. This property is required for any distribution to accurately estimate the distribution of the preposterior mean, as seen in Figure 9.

Given the linear transformation in (20), it is possible to write down our approximation for the density of the preposterior mean directly as a function of the density of the prior $p(\mathsf{NB}_t^{\boldsymbol{\theta}})$

$$p(\mu_t^{\boldsymbol{X}}) \approx \frac{1}{a}\, p\left(\frac{\mu_t^{\boldsymbol{X}} - b}{a}\right).$$

Note that this density does not approximate the function $\mu_t^{\boldsymbol{X}}$ itself but the *density* $p(\mu_t^{\boldsymbol{X}})$.

As samples from the prior of the net benefit are available in the form of PSA simulations, approximating the density of the preposterior mean is simply a matter of estimating the constants a and b. However, as the prior mean and variance can be calculated from the available prior samples, the approximation of these two constants reduces to estimating the expected variance of the preposterior mean (§3.4). Additionally, note that these formulæ are simply based on means and variances and are therefore very simple to calculate, while the proposed transformation of the prior is also very simple.

3.3.2.2 *Why does the moment matching estimation work?*

Before considering the estimation method for the variance of the preposterior mean, we provide some justification for this moment matching method. To begin, consider two special cases; at one extreme, assume that $\boldsymbol{X}$ is independent of the underlying model parameters: $p(\boldsymbol{X} \mid \boldsymbol{\theta}) = p(\boldsymbol{X})$. Evidently, this setting would never occur as decision makers only consider data collection that would aid the decision making process. Nevertheless, if the sample is independent of the model parameters then the distribution of the preposterior mean is a point mass at the prior mean

$$\mathsf{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right] = \mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right],$$

by the condition of independence.

Using the definition for a and b from (20), note that

$$a = \sqrt{\frac{\mathsf{Var}\left[\mathsf{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right]\right]}{\mathsf{Var}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right]}} = 0$$

and

$$b = \mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right](1-a) = \mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right],$$

which means that the approximation for the density of the preposterior mean is also equal to the prior mean and therefore exact for all model structures when the sample and the model parameters are independent. While, practically, this is a relatively unimportant result, it does indicate that our approximation is roughly accurate when the variance of the preposterior mean is small.

At the other end of the scale, it is possible to show that our approximation using moment matching is exact when the sample is deterministically linked to the model parameters, i.e. $\boldsymbol{X} = h(\boldsymbol{\theta})$ for some $h(\cdot)$. In this setting, the *conditional* mean for the net benefit is equal to the known-distribution net benefit since, if the value for $\boldsymbol{X}$ is known, then the exact $\mathsf{NB}_t^{\boldsymbol{\theta}}$ value

is also known. In a similar manner to above, it can be shown that $a = 1$ and $b = 0$ so the approximation for the density of the preposterior mean is equal to $\mathsf{NB}_t^{\boldsymbol{\theta}}$ which, again, is the exact distribution for the preposterior mean.

Therefore, as the variance of the preposterior mean increases, the approximation becomes exact. This has a practical implication since, provided the posterior is consistent, this approximation is accurate for large sample sizes. This is because, as the sample size N increases, the distribution of the preposterior mean reflects the prior more closely (Figure 9) as the sample contains more information about the underlying values for $\boldsymbol{\theta}$.

To extend these ideas, this approximation is accurate for moderate N when the posterior mean net benefit is a weighted average between the prior mean and a data summary

$$\mathsf{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right] = c\ \mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right] + d\ g(\boldsymbol{X}),$$

where c and d are constants and $g(\cdot)$ is an arbitrarily complex function of the data which must have a similar density to $\mathsf{NB}_t^{\boldsymbol{\theta}}$.

For all conjugate settings in the exponential family, the posterior mean can be written as weighted average of the prior mean and a data summary [39]. Therefore, it is sufficient to consider whether the prior predictive distribution of the data summary has a similar distribution to the prior. In the simplest setting, it is possible to demonstrate that this is true in the normal-normal setting, as seen in §3.2.2.2, and therefore, our approximation will be accurate when the prior for the net benefit is approximately normal, coupled with an approximately normal distribution for $g(\boldsymbol{X})$. In §3.5.3, we demonstrate that the approximation can give biased estimates in non-normal settings. However, the bias is minimal for realistic sample sizes and decreases further as the sample size N increases since the preposterior distribution approaches the prior and the variance of the preposterior mean tends to the prior variance.

3.4 APPROXIMATING THE PREPOSTERIOR MEAN DISTRIBUTION BY SIMULATION

As previously discussed, using the moment matching methodology reduces approximating the distribution of the preposterior mean to estimating the constants a and b from §3.3.2, under the assumption (usually true, in the context of health economic evaluation) that simulations from the distribution for $\mathsf{NB}_t^{\boldsymbol{\theta}}$ under current information are available. These constants are based on the mean and variance of the prior for the net benefit (i.e. prior to the future sample $\boldsymbol{X}$), and the expected posterior variance over all possible samples $\boldsymbol{X}$. Therefore, the following section is concerned with estimating the expected posterior variance using a *small* number of posterior samples. This reduces the number of posterior samples needed compared to the nested Monte Carlo simulation and therefore reduces the computational time required to approximate the distribution of the preposterior mean.

3.4.1 *Estimating the variance of the preposterior mean*

To begin, it may seem that estimating the expected posterior variance over different possible samples of $\boldsymbol{X}$ by Monte Carlo simulation would not save computational time compared to estimating the distribution of the preposterior mean by finding the posterior mean for different samples. However, in general, the posterior variance is relatively *stable* implying that the posterior variance changes relatively little across the different future samples $\boldsymbol{X}$ compared to the posterior mean. This stability is most extreme in the normal-normal conjugate setting where the posterior variance is independent of the posterior mean and dependent simply on the variance of the sample $\boldsymbol{X}$, not its location. Therefore, the posterior variance is the same for each future sample $\boldsymbol{X}$ for a fixed sample size N, implying that only one posterior sample is required to estimate the *expected* posterior variance which is then used to estimate the constant b (20).

With substantial departures from normality, the posterior variance is no longer independent of the location of the samples, but, as shown in §3.6, a small number of posterior samples (around 20-50) can be used to estimate the expected posterior variance, even in highly non-normal settings. This is because quadrature can be employed to reduce the number of hypothetical posterior variance estimates that are needed for this estimation.

Specifically, quadrature is employed to calculate

$$\mathsf{E}_{\boldsymbol{X}}\left[\mathsf{Var}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right]\right] = \mathsf{E}_{\boldsymbol{\theta}}\left[\mathsf{E}_{\boldsymbol{X}|\boldsymbol{\theta}}\left[\mathsf{Var}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t^{\boldsymbol{\theta}}\right]\right]\right],$$

where the two outer expectations on the RHS allow us to take an expectation over the prior predictive distribution without direct sampling. We use quadrature to estimate the outer expectation with respect to $\boldsymbol{\theta}$ and Monte Carlo simulation across the different quadrature points for the inner expectation with respect to $\boldsymbol{X} \mid \boldsymbol{\theta}$.

In most data collection exercises, $\boldsymbol{X}$ will only be directly conditional on a small number of parameters as in general researchers are only directly interested in 1 or 2 parameters when designing a trial. Therefore, the outer expectation, with respect to $\boldsymbol{\theta}$ above, will typically be over a uni- or bi-variate vector, in which case quadrature is simple to implement. In §3.4.3, we further the discussion of situations where the sample is only dependent on a sub set of the model parameters. More broadly, it is possible to sample from the prior-predictive distribution of the data although it is then more challenging to determine how to implement quadrature in this setting §4.4.2.

In general, we suggest that each of the parameters that are directly informed by $\boldsymbol{X}$ should be reordered so they are in ascending order. The $\frac{q}{Q+1}^{\text{th}}$, for $q = 1, \ldots, Q$ values in these reordered lists should then be selected, giving the sample quantiles for each of the parameters updated by $\boldsymbol{X}$. These sample quantiles are used to give good coverage over all possible values of $\boldsymbol{\theta}$. Each future sample $\boldsymbol{X}$ is simulated conditional the sample quantiles for each of the elements in $\boldsymbol{\phi}$. These samples are then used to update the posterior and calculate the variance. As MC simulations are used to estimate the inner expectation, we recommend in excess of 30 simulations to avoid the dependence on specific samples. Notice here that there

is a clear trade-off between accuracy of the estimate for the variance of the preposterior mean and the computational time required to obtain this estimate, we return to this consideration in §3.6. The advantage of this method is that good coverage is achieved with relatively low computational cost compared to standard MC simulation. §3.8.2 discusses a brief extension to this method to ensure greater coverage.

3.4.2 *Calculating the EVSI for a specific set of treatment options*

Thus far, we demonstrated how to estimate the distribution of the preposterior mean for the net benefit. However, to calculate the EVSI, we need to compute the joint distribution of the preposterior means across the different treatments to find the dominant treatment. In general, this requires the estimation of the posterior variance-covariance matrix for the net benefits for the different treatments.

While this adds little theoretical complexity, this EVSI estimation method is more stable if we work directly with the incremental net benefit (INB); defined as the difference between two treatment options, e.g. $\text{INB}_t^{\boldsymbol{\theta}} = \text{NB}_t^{\boldsymbol{\theta}} - \text{NB}_0^{\boldsymbol{\theta}}$, $t = 1, \ldots, T$. In dual decision settings (i.e. $T = 1$), the optimal treatment is found by comparing $\text{INB}_1^{\boldsymbol{\theta}}$ with 0; if $\text{INB}^{\boldsymbol{\theta}} > 0$ then treatment 1 ($t = 1$) is optimal and if it is negative $t = 0$ is optimal. The distribution of the preposterior mean $\text{INB}_1^{\boldsymbol{\theta}}$ is then estimated using moment matching, and the EVSI is calculated using

$$\text{E}_X \left[\max \left\{ 0, \text{E}_{\boldsymbol{\theta}|X} \left[\text{INB}_1^{\boldsymbol{\theta}} \right] \right\} \right] - \max \left\{ 0, \text{E}_{\boldsymbol{\theta}} \left[\text{INB}_1^{\boldsymbol{\theta}} \right] \right\},$$

where $\mu^X = \text{E}_{\boldsymbol{\theta}|X} \left[\text{INB}_1^{\boldsymbol{\theta}} \right]$ is only based on scalar mean and variance values rather than a mean vector and a variance-covariance matrix.

When more than two options are under consideration (multi-decision setting), it is also preferable to work with the INB as it reduces the size of the variance-covariance matrix, i.e. the EVSI can be calculated based on the variance-covariance matrix for the distribution of $\text{INB}_1^{\boldsymbol{\theta}}$ and $\text{INB}_2^{\boldsymbol{\theta}}$ rather than $\text{NB}_0^{\boldsymbol{\theta}}$, $\text{NB}_1^{\boldsymbol{\theta}}$ and $\text{NB}_2^{\boldsymbol{\theta}}$. This reduces the number of parameters that need to be estimated as there are only 3 unique elements in the variance-covariance matrix rather than 6.

3.4.3 *Nuisance Parameters*

In realistic health economic models, it is very unlikely that the proposed data collection exercise will directly inform all the model parameters $\boldsymbol{\theta}$. For example, a clinical trial may only give information about the drug effectiveness and not the societal costs of the disease. In fact, the future study should be designed to target only those parameters that have value, as shown by the EVPPI analysis. Therefore, we will be considering studies that target uncertainty in a subset of the underlying model parameters $\boldsymbol{\phi} \subset \boldsymbol{\theta}$.

In general, there is no guarantee that the shape of the prior distribution for $\text{INB}_1^{\boldsymbol{\theta}}$ conditional on all the model parameters $\boldsymbol{\theta}$ will be the same as the distribution of $\text{INB}_1^{\boldsymbol{\theta}}$ conditional on $\boldsymbol{\phi}$.

To demonstrate the phenomenon, we introduce a simple two parameter model: $\phi \sim Be(1,4)$ and $\psi \sim N(-0.5,1)$ where $\text{NB}_0^{\boldsymbol{\theta}} = 10\,000\psi - 4\,000$; $\text{NB}_1^{\boldsymbol{\theta}} = 10\,000\phi - 6\,500$; and $\text{INB}_1^{\boldsymbol{\theta}} = 10\,000(\phi - \psi) - 2\,500$.

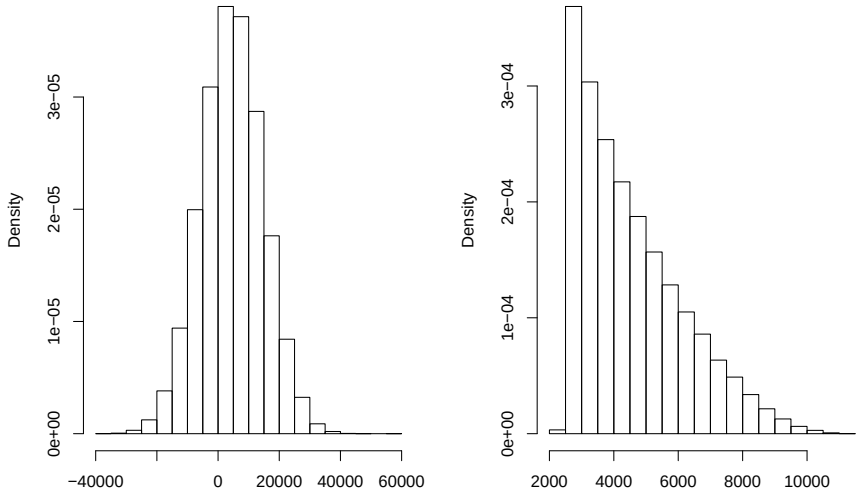


Figure 10: The distribution of the incremental net benefit conditional on both the model parameters $\boldsymbol{\theta}$ (LHS) and conditional on the parameter of interest ϕ (RHS) for the simple two-parameter model.

Figure 10 (LHS) shows the prior for $\text{INB}_1^{\boldsymbol{\theta}}$ which has a similar shape to a normal distribution — particularly in the tails. However, as $\text{INB}_1^{\boldsymbol{\theta}}$ is linear in ϕ, the distribution of $\text{INB}_1^{\boldsymbol{\theta}}$ conditional on ϕ would be a shifted and scaled Beta distribution which has a very different shape, as seen in Figure 10 (RHS). Therefore, if the prior for $\text{INB}_1^{\boldsymbol{\theta}}$ was used to approximate the distribution of the preposterior mean when $\boldsymbol{X}$ only updates information about ϕ, the mean and variance would be correct but the shape would be approximately normal which would lead to inaccurate estimates for the EVSI.

To identify a more appropriate shape for the distribution of the preposterior mean, the uncertainty due to the nuisance parameters $\boldsymbol{\psi}$ should be marginalised out. This means that the nuisance parameters do not impact the shape of the approximate distribution of the preposterior mean:

$$\int_{\boldsymbol{\Psi}} \text{INB}(\boldsymbol{\phi}, \boldsymbol{\psi}) p(\boldsymbol{\phi}, \boldsymbol{\psi}) d\boldsymbol{\psi} = \text{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}[\text{INB}_t^{\boldsymbol{\theta}}] = \text{INB}_t^{\boldsymbol{\phi}}. \tag{21}$$

Note, however, that this marginalisation is exactly the calculation required to estimate the EVPPI and should therefore already be available when trying to calculate the EVSI [117]. However, if these values are not available then the methods in Chapter 2 can be used to efficiently marginalise out uncertainty due to $\boldsymbol{\phi}$.

Note that, to rescale the sample of the $\text{INB}_t^{\boldsymbol{\phi}}$, rather than $\text{INB}_t^{\boldsymbol{\theta}}$, the constant a from equation (20) becomes:

$$a = \frac{\sigma}{\sqrt{\text{Var}_{\boldsymbol{\phi}}\text{INB}_t^{\boldsymbol{\phi}}}},$$

where σ, the standard deviation of the preposterior mean, must still be calculated as a function of the variance of $\text{INB}_t^{\boldsymbol{\theta}}$ — not the variance of $\text{INB}_t^{\boldsymbol{\phi}}$ — and the mean posterior variance, as outlined in §3.4.1.

3.4.4 *The Moment Matching algorithm for EVSI estimation*

Finally, to clarify the Moment Matching method, we briefly present the algorithm in the dual-decision setting. The extension to multi-decision is trivial but slightly complicates the notation and so is revisited in §4.4.3.

1. Obtain the PSA simulations $\boldsymbol{\theta}_s$ and $\text{INB}_1^{\boldsymbol{\theta}_s}$ for $s = 1, \ldots, S$.
2. From these simulations, estimate the mean and variance of $\text{INB}_1^{\boldsymbol{\theta}}$, denoted μ and σ^2 respectively.
3. Determine the parameter(s) $\boldsymbol{\phi}$ that will be updated by proposed data collection exercise.
4. Estimate the $\text{INB}_1^{\boldsymbol{\phi}_s}$, the expectation of $\text{INB}_1^{\boldsymbol{\theta}}$ conditional on $\boldsymbol{\phi}_s$ only, for $s = 1, \ldots, S$, and its variance $\sigma^2_{\boldsymbol{\phi}}$.
5. Find Q quantiles of the PSA simulations separately for each element of $\boldsymbol{\phi}$. We denote the vector that contains the q-th quantile for each element of $\boldsymbol{\phi}$, $\boldsymbol{\phi}_q$ for $q = 1, \ldots, Q$.
6. For each $\boldsymbol{\phi}_q$, sample a potential future dataset $\boldsymbol{X}_q$ from $p(\boldsymbol{X} \mid \boldsymbol{\phi}_q)$.
7. For each future dataset $\boldsymbol{X}_q$, update the posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{X}_q)$ and calculate $\text{INB}_1^{\boldsymbol{\theta}}$. From this, calculate the variance of $\text{INB}_1^{\boldsymbol{\theta}}$, denoted σ^2_q.
8. Estimate the variance of $\mu_1^{\boldsymbol{X}}$ with the following formula

$$\sigma^2_{\boldsymbol{X}} \approx \sigma^2 - \frac{1}{Q}\sum_{q=1}^{Q}\sigma^2_q.$$

9. Rescale $\text{INB}_1^{\boldsymbol{\phi}_s}$ for $s = 1, \ldots, S$:

$$\text{INB}^*_{1,s} = \frac{\text{INB}_1^{\boldsymbol{\phi}_s} - \mu}{\sigma_{\boldsymbol{\phi}}}\sigma_{\boldsymbol{X}} + \mu.$$

10. Estimate the EVSI using $\text{INB}^*_{1,s}$:

$$\text{EVSI} \approx \frac{1}{S}\sum_{s=1}^{S}\max\left\{0, \text{INB}^*_{1,s}\right\} - \max\left\{0, \mu\right\}$$

3.5 EXAMPLES FOR THE EVSI ESTIMATION PROCEDURE

To test this approximation for the distribution of the preposterior mean, we now estimate the EVSI in several different examples. This section begins with two "toy" examples that exploit conjugacy to find both analytic results and computationally efficient algorithms to calculate both the true distribution for the preposterior mean and the true EVSI. These examples are used to explore some difficulties associated with estimating the EVSI using this methodology and demonstrate situations when it is suitable. Firstly, the Beta-Binomial example from §3.2.2 is extended to demonstrate the difficulties associated with using a continuous approximation for discrete samples. Exponential-Gamma conjugacy is then exploited to demonstrate this methodology where the data summary does not have the same distribution as the prior.

Following this, a more realistic health economic model is presented to explore the full capabilities of this methodology. This is based on a decision tree model developed in [1] and explores the EVSI estimation procedure in the presence of nuisance parameters $\boldsymbol{\phi}$. The new methodology presented here is also contrasted with EVSI estimation methods based on sufficient statistics [112] and MC simulation.

Finally, we investigate the variance estimation procedure by considering the performance of the estimator of the variance of the preposterior mean in a highly non-normal setting. We also use a chemotherapy example [4] to explore the optimal way to choose Q — the number of quadrature points for the variance estimation.

3.5.1 *Discrete Samples with Beta-Binomial Conjugacy*

Revisiting the first example presented in §3.2.2, the parameter θ is modelled using a vague Beta prior $\theta \sim Beta(1,1)$ and the data have a binomial distribution $X \mid \theta \sim Bin(N,\theta)$. The two net benefit functions are then $\text{NB}_0^\theta = 0$ and $\text{NB}_1^\theta = k\theta - c$, so $\text{INB}_1^\theta = k\theta - c$.

In this setting, the approximation of the distribution of the preposterior mean could be poor, as the data collection exercise is discrete, implying that the true distribution of the preposterior mean is also discrete, while the prior for INB_1^θ is continuous. This phenomenon can be seen most clearly when the binomial sample size $N = 1$ and there are two equally likely possible samples, $X = 0$ and $X = 1$. This implies that there are two equally likely possible preposterior means; $\mu_1^0 = \frac{k}{3} - c$ or $\mu_1^1 = \frac{2k}{3} - c$. Clearly, this distribution will never be well approximated by a shifted and rescaled beta distribution.

To investigate when such a continuous approximation is suitably accurate, we estimate the EVSI for different possible sample sizes N. Due to the conjugate structure, it is possible to calculate both the EVSI and the variance of the preposterior mean analytically. The true

variance of the preposterior mean is then used to calculate a and b from §3.3.2. These are then used to approximate the EVSI by shifting and rescaling the prior for the net benefit:

$$\text{EVSI} \approx \frac{1}{10\,000} \sum_{s=1}^{10\,000} \max\left\{0, a\,\text{INB}_1^{\theta_s} + b\right\},$$

where θ_s is the s-th simulated value from the prior for θ (in this case, we use a *simulation* size of $10\,000$).

While the variance of the preposterior mean is available analytically, the estimator for the EVSI is still subject to variability due to the specific prior simulation of θ. Therefore, $10\,000$ different simulations of size $10\,000$ were taken from the prior for θ and used to calculate the EVSI with the moment matching method to find the sampling distribution of the EVSI estimator. This distribution quantifies the uncertainty in the EVSI estimate that arises from using simulation to estimate the distribution of INB_1 and using moment matching to estimate μ_1^X, as the variance is estimated analytically, this distribution excludes uncertainty in the variance estimation procedure. Therefore, if moment matching is accurate, these distributions should be centred on the true analytic value for the EVSI with little variance. This uncertainty would be difficult to estimate in practice.

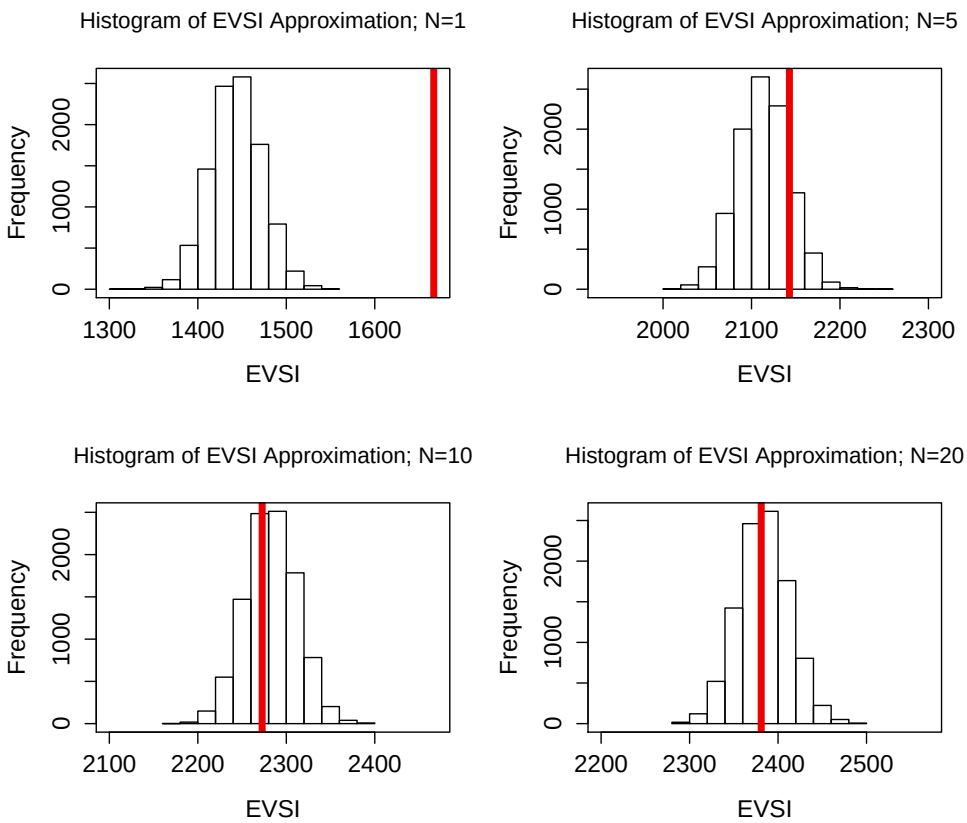


Figure 11: The distribution of the EVSI estimator over 10 000 different simulations from the prior of θ for 4 different sample sizes for X for the Beta-Binomial conjugate model. The red line represents the analytical value of the EVSI.

Figure 11 shows the sampling distribution of the EVSI estimator for different sample sizes for X, with a red line marking the sample specific EVSI for each sample size N. The top LHS shows that the EVSI estimator for $N = 1$ has a significant downward bias as the sampling distribution does not include the true EVSI value of 1667. This clearly indicates that the weighted prior distribution is not a suitable approximation for the distribution of the preposterior mean in this setting, as expected. However, as the sample size increases, the bias decreases and for a sample size of only 10 it is negligible, as the distribution of the EVSI is centred at the true value.

This analysis indicates that even if the distribution of the preposterior mean is discrete, it can be well approximated when the sample size for X is sufficiently large, > 10 in this example. In general, if the estimation of the EVSI is the primary interest, then a smaller sample size may be permitted, as we rarely require that the EVSI is estimated to a high degree of accuracy. This is because the EVSI is compared with trial costs which are rarely known with certainty. Additionally, the EVSI is based on incorrect model assumptions meaning that even if the EVSI is highly accurate given the model, it will be approximate when applied in practice.

3.5.2 *Non-linear mean function with Exponential-Gamma conjugacy*

We now revisit the second example in §3.2.2 where a Gamma prior is assumed for the parameter of interest $\theta \sim Gamma(\alpha, \beta)$, the data collection exercise is assumed to be N independent observations from an exponential distribution $X_j \sim Exp(\theta)$, $j = 1, \ldots, N$ and the two net benefit functions are:

$$\text{NB}_0^\theta = c_0 \quad \text{and} \quad \text{NB}_1^\theta = k\theta - c_1.$$

Throughout §3.5.2, we present the results for $\alpha = 5$, $\beta = 1$, $k = 200$ and $c_0 = 900$ and $c_1 = 100$ as in Figure 9.

Using Gamma-Exponential conjugacy, it is trivial to show that the preposterior mean is equal to

$$\mu_1^{\boldsymbol{X}} = \text{E}_{\theta|\boldsymbol{X}}\left[\text{INB}_1^\theta\right] = k\frac{\alpha + N}{\beta + \sum_{i=1}^N X_i} - c_1 - c_0,$$

which means that both the variance of the preposterior mean and the EVSI can be found analytically, as for the previous example. Therefore, any difficulties in estimating the EVSI are because the weighted prior distribution is not a suitable approximation to the distribution of the preposterior mean. As this is a conjugate model, this misspecification is because the data summary does not have the same distribution as the prior.

Figure 12 shows the sampling distribution of the EVSI values, over different prior samples, for different values of N. The uncertainty represented in this sampling distribution is the same as for the Beta-Binomial example §3.5.1. In this example, moment matching with the prior gives a biased EVSI estimate for small samples. However, this bias is at most 4% of the total EVSI value, meaning that it can still be used, as the estimate is sufficiently accurate for decision making. It is recommended, however, to remember that the EVSI estimate is slightly

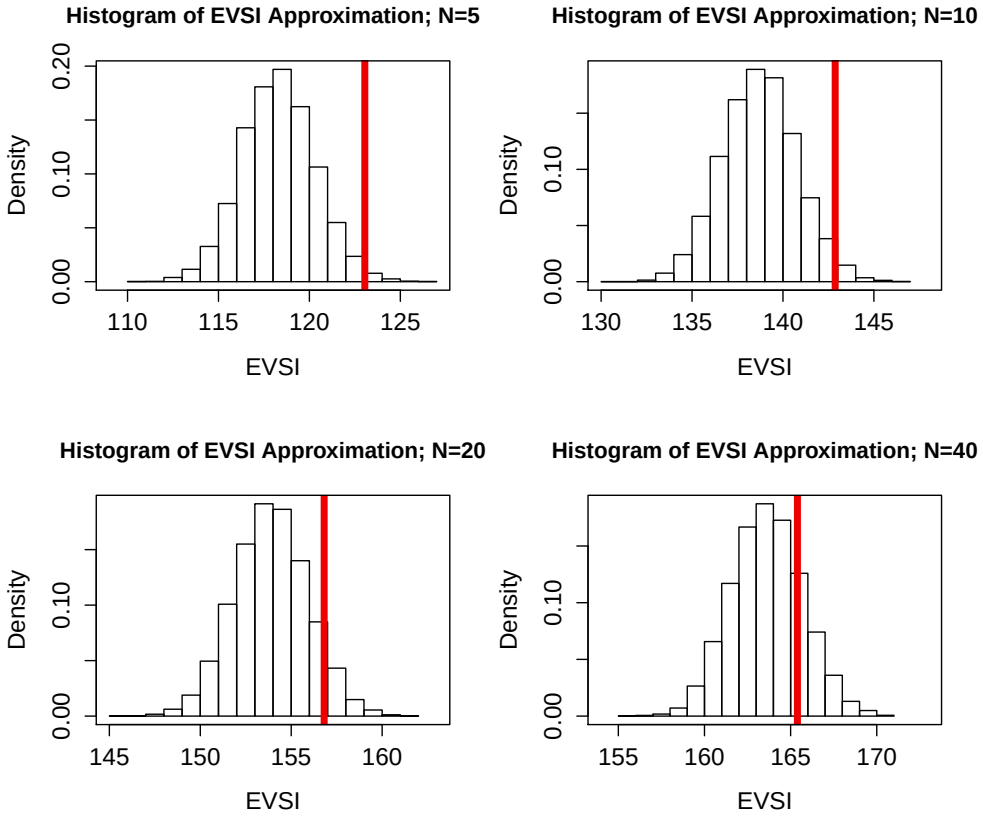


Figure 12: The distribution of the EVSI estimator over $10\,000$ different simulations from the prior of θ for 4 different samples sizes for $\boldsymbol{X}$ using the Exponential-Gamma conjugate model. The red line represents the analytic value of the EVSI.

biased for small sample sizes and therefore care should be taken interpreting the EVSI for these small samples. As N increases, this bias becomes negligible as the distribution of the preposterior mean tends exactly to the prior as the sample size increases, see Figure 9.

3.5.3 *Case Study: Ades et al. Decision Tree Model*

To demonstrate the effectiveness of our methodology in a more practical scenario, we use a decision tree model developed in Ades *et al.* [1]. This model has two treatment options, a standard of care and a new treatment, aimed at avoiding a critical event. This critical event leads to a reduction in QALYs for the remainder of the patient's life. The new treatment reduces the probability of the critical event but the patient may also experience side effects which give a short term reduction in QALYs along with a direct cost of additional treatment. The model has 11 parameters, of which 4 are subject to uncertainty which is then modelled using 4 mutually independent distributions; a complete model description is given in Ades *et al.* [1] or Strong *et al.* [112].

For this case study, we consider four different data collection exercises, the first three have been tackled by Ades *et al.* and the fourth investigates the moment matching method when $\boldsymbol{\phi}$ is not unidimensional;

1. To reduce uncertainty in the probability of side effects for the new treatment ϕ_1, 60 patients are given the new treatment and the number who suffer from side effects is recorded.

2. To reduce uncertainty in the quality of life after the critical event ϕ_2, the quality of life for 100 patients who experienced the event is recorded.

3. To reduce uncertainty in the odds ratio of the effectiveness of the two treatments ϕ_3, a randomised control trial with 200 patients on each arm is undertaken.

4. To reduce uncertainty in the probability of the critical events in both treatment arms ϕ_3^C, ϕ_3^T, the same randomised control trial is undertaken but the analysis informs these two probabilities directly.

For a full description of the distributional assumptions for these studies, particularly the difference between study 3 and study 4, see Strong *et al.* [112].

3.5.3.1 *Computations for the Ades et al. Model*

To calculate the EVSI using the moment matching methodology, $1\,000\,000$ simulations were taken for the 4 stochastic model parameters. These are then combined with the other seven parameters to calculate $1\,000\,000$ simulations for $\text{INB}_1^{\boldsymbol{\theta}}$ under current information. These simulations for $\text{INB}_1^{\boldsymbol{\theta}}$ were used to find $\text{INB}_1^{\boldsymbol{\phi}}$, the fitted $\text{INB}_1^{\boldsymbol{\theta}}$, using GAM regression [55] obtained with the `gam` function from the `mgcv` package [126] in `R`. The simulations were also used to find the mean and variance of $\text{INB}_1^{\boldsymbol{\theta}}$.

To estimate the preposterior variance, the expected posterior variance was estimated using MCMC procedures with Q, the number of quadrature points, equal to 30. This was achieved using `JAGS` through `R` [92] with $10\,000$ simulations from each posterior distribution and $1\,000$ simulations used as burn-in. This means that in total $1\,330\,000$ simulations were used to estimate the EVSI in this example, although note that the PSA simulations were reused for each EVSI calculation. Fewer simulations could be used but as this example has a small computational cost it was possible to use this number of simulations to improve the accuracy of the methods. The computational time required to estimate the posterior variance with this number of simulations was between 3.8-6.1 seconds.

To assess the accuracy of our method, the expected posterior variance was estimated using the above procedure $1\,000$ times for each trial. Each of these estimates for the expected posterior variance was then used to approximate the EVSI using moment matching to give a distribution for the EVSI estimate obtained using moment matching. Therefore, the sampling distributions given in §3.5.3.2 represent uncertainty in the variance estimation procedure but do not take into account MC error arising from the prior simulations. This represents uncertainty in Q but not in the size of the PSA simulations S. In a standard analysis, our method

would produce a point estimate for the EVSI rather than these distributions and it would be impossible for a user to access this distribution.

To determine the accuracy of our method, we compared with the computationally intensive two-step nested Monte Carlo procedure [12] and the Strong *et al.* method [112] based on sufficient statistics and non-parametric regression which is an accurate and efficient estimation method for models where the data can be summarised using a low dimensional sufficient statistic. For the experiments targeting ϕ_1, ϕ_2 and ϕ_3, the comparator values are taken directly from the Strong *et al.* paper where the estimates are based on 10^{10} and 10^6 simulations respectively. For the two parameter EVSI estimate, the results for these two methods were obtained using the same number of simulations. The computational times to obtain these estimates were 8.4 seconds for the Strong *et al.* method and $207\,595$ seconds (approximately 2.4 days) for the nested Monte Carlo method.

3.5.3.2 *EVSI estimates for the Ades et al. Model*

Figure 13 plots the sampling distribution of the EVSI estimate, over the different estimates of the expected posterior variance, obtained using our moment matching method for the four alternative study designs. The solid line gives the value of the estimate obtained by nested Monte Carlo (which can be considered as the "truth"), the dashed line is the estimate obtained using the Strong *et al.* method and the dotted line is the average moment matching estimate. Evidently, the moment matching method is in line with these two alternative estimation methods for the studies considered, although the two-parameter estimation has a slight upward bias. Nonetheless, the true EVSI value is well within the 90% interval, represented by the solid black line on the axis in Figure 13.

Figure 13 also demonstrates that the EVSI estimate calculated using moment matching is less accurate for experiments with smaller EVSI — as the EVSI gets smaller, the 90% intervals get wider. This is because the estimate is based on the difference between the prior variance and the expected posterior variance. When the EVSI is small then this difference is also small and therefore the posterior and prior variance needs to be estimated with greater precision because the difference can be greatly affected by the Monte Carlo error in the variance estimation. It is important to note that the accuracy demonstrated in Figure 13 depends on accurately estimating the prior variance. Therefore, our moment matching method should not be used if the initial PSA simulation size is very small.

3.6 CALCULATING THE VARIANCE OF THE PREPOSTERIOR MEAN

As stated in §3.4.1, the estimation for the variance of the preposterior mean is highly effective when the prior for the $\text{INB}_t^{\boldsymbol{\theta}}$ is roughly normal. Therefore, to test this procedure we assume a highly non-normal prior for the $\text{INB}_t^{\boldsymbol{\theta}}$. Counterintuitively, it is possible to exploit normal-normal conjugacy to carry out this investigation by setting $\text{INB}_1^{\theta} = \theta^2 - 5$ where θ is normal *a priori* with mean 0 and precision 0.2: $\theta \sim N(0,5)$. The data collection is then assumed to be 10 independent observations $X_j \sim N(\theta,1)$ for $j = 1,\ldots,10$. Conjugacy can now be used to

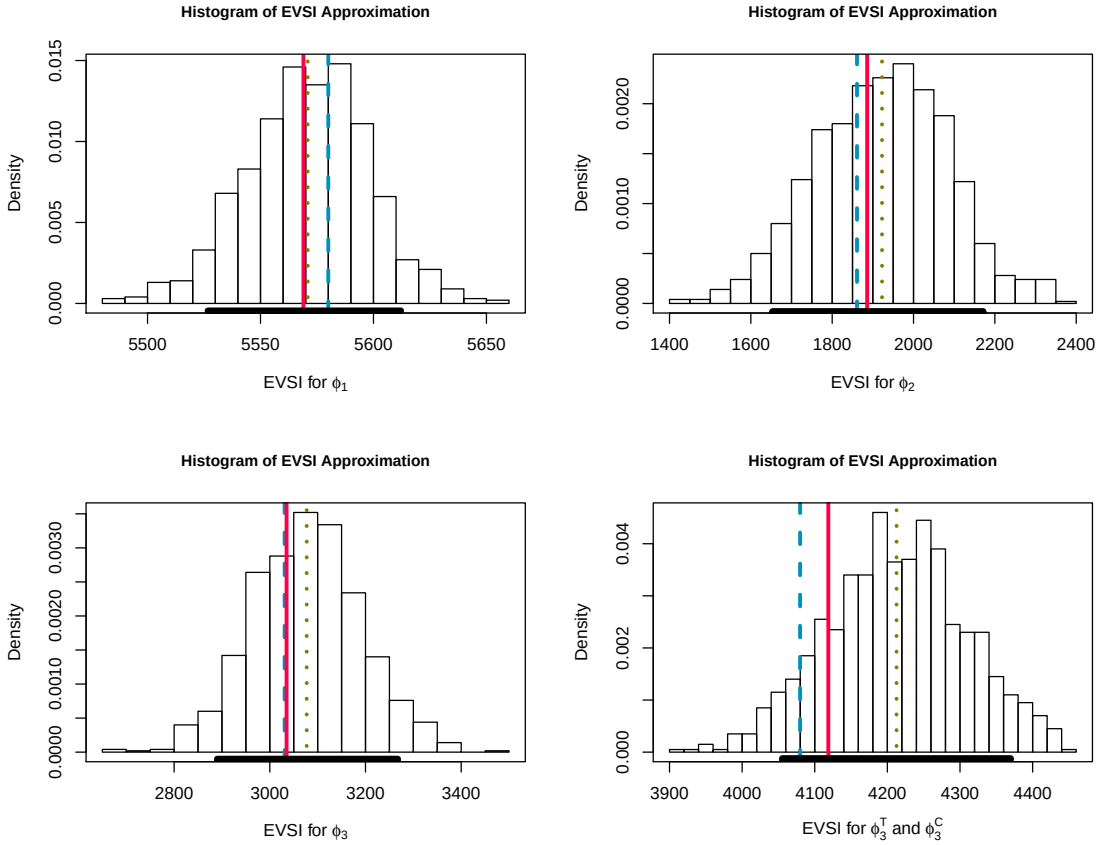


Figure 13: The "sampling distribution" of the EVSI conditional on the distribution over the different estimates for the variance of the preposterior mean for ϕ_1, ϕ_2 and ϕ_3 for the Ades *et al.* example [1]. The solid red line represents the EVSI calculated using Monte Carlo methods and 10^{10} simulations. The dashed blue line represents the EVSI estimate obtained using the Strong *et al.* method with 10^6 simulations. The dotted brown line represents the average EVSI estimate obtained using moment matching and the solid black horizontal line represents the 90% interval for this estimation method. The comparator methods (MC and Strong *et al.*) are taken from Strong *et al.* [112] for the first three graphics.

calculate a value for the posterior for θ efficiently, while inducing a highly non-normal prior for the INB_1^θ.

Using conjugacy, it is possible to estimate both the EVSI and the variance of the preposterior mean cheaply using Monte Carlo methods [1]. Therefore, using $10\,000$ samples from the prior for the INB_1^θ, the EVSI is estimated as 2.00 and the variance of the preposterior mean is 35.20. As these estimates are obtained using simulation methods, there is some uncertainty surrounding these estimates of the "true" variance of the preposterior mean.

The estimation method for the variance of the preposterior mean requires Q quadrature points spaced throughout the domain of θ. Practically these are taken as the Q quantiles for θ, i.e. the $S\frac{q}{Q+1}$–th θ values in an ordered sample, with $q = 1, \ldots, Q$.

Figure 14 shows the average estimate, over 500 simulations, for the preposterior variance for increasing values of Q up to $Q = 100$ — this means that $1\,000$ simulations were taken from

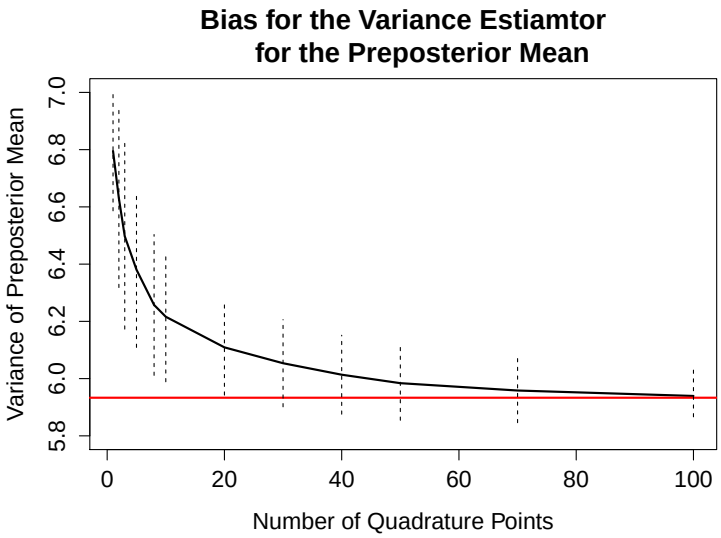


Figure 14: The estimate of the variance of the preposterior mean for increasing numbers of quadrature points. The red line gives the variance of the preposterior mean calculated using all the samples in the prior for θ. The dashed lines are the standard errors for the estimates of the variance of the preposterior mean.

100 different posterior distributions 500 times. The red line in Figure 14 shows the estimated value of the variance of the preposterior distribution calculated using the method from [1]. The dashed lines indicate plus or minus one standard deviation from the mean estimate of the variance of the preposterior distribution for the different sample sizes. The standard deviation, in this setting, represents the uncertainty from limited Q and limited posterior samples. It does not include uncertainty due to the MC estimation of the variance of the prior. Again, it would not be possible to calculate this standard error using one estimation method.

In general, our estimation method for the variance of the preposterior mean produces biased estimates for small numbers of quadrature points. However, once the number of quadrature points exceeds 30, the true variance is within one standard deviation of the average estimate of the variance of the preposterior mean. The standard deviation of this estimate also decreases as the number of quadrature points increases.

Table 6 shows the EVSI estimate and its bias when we use the estimated variance of the preposterior mean given in Figure 14 and our moment matching methodology. Notice that while all the EVSI values are over-estimated the percentage bias drops below 0.02 for more than 30 quadrature points. Therefore, it seems that a relatively small number of quadrature points are reasonable for estimating the variance of the preposterior mean, even in significantly non-normal settings.

Although Table 6 demonstrates that the variance estimate will become more accurate as the number of quadrature points increases, using more quadrature points does involve a greater computational burden. Thus, estimating the EVSI using this method requires a trade-off between computational time and accuracy. In this sense, we need to determine an optimal value of Q that allows us to trade off these two considerations. To investigate potential values

Number of simulations	1	2	3	5	8	10	∞
Estimate of EVSI	2.30	2.24	2.20	2.16	2.12	2.10	2.00
Percentage Bias	0.15	0.12	0.10	0.08	0.06	0.05	0.00
Number of simulations	20	30	40	50	75	100	∞
Estimate of EVSI	2.06	2.05	2.03	2.02	2.01	2.01	2.00
Percentage Bias	0.03	0.02	0.02	0.01	0.01	0.01	0.00

Table 6: The EVSI estimate for different numbers of posterior samples using the moment matching method.

for Q, a simple, but relatively realistic health economic model for a new chemotherapy drug is used.

3.6.1 *A new chemotherapy drug*

This model assesses the cost-effectiveness of a new chemotherapy drug for the treatment of cancer against the standard of care [4]. After the treatments are administered, the patients may experience haematological (blood related) side effects. If these side effects occur then the patient either requires hospital admission or ambulatory care, depending on the severity of the side effects. The new drug is assumed to reduce the likelihood of having side effects and therefore reduces the cost of treating the side effects while increasing the quality of life for the patients.

There are 10 model inputs; c_H and c_A, the cost of hospital and ambulatory care respectively; A_t, H_t and SE_t for $t = 0, 1$, the number of patients requiring hospital care, ambulatory care or experiencing side effects for the two different treatments; π_t, the probability of experiencing side effects for the two treatment options and γ, the probability of requiring hospital care once side effects have been experienced. The effectiveness measure in this is example is taken to be the total number of people who do not experience side effects and the costs are equal to the total cost of treatment the drug related side-effects. Finally, to compute the VoI measures, the willingness-to-pay parameter, representing the amount of money the payer is willing to pay to avoid one person being affected by side effects, is taken as £$6\,000$.

Before calculating the EVSI, we perform a full VoI analysis [117] to determine whether the decision uncertainty is sufficiently high to warrant additional research. Firstly, the EVPI is calculated as £$104\,549$ which is sufficiently high that a future study to reduce decision uncertainty would likely have value. Furthermore, to determine where to direct future research, we consider the single-parameter EVPPI for all the underlying model parameters, given in Table 7. This analysis indicates that the parameter driving decision uncertainty is the probability of experiencing side effects on the treatment arm, π_1. Therefore, a study is designed to learn more information about π_1.

Parameter	A_0	A_1	H_0	H_1	SE_0
EVPPI	16350	30320	30740	18480	35750
SE_1	c_a	c_h	π_0	π_1	γ
35250	0	4780	22930	24660	6850

Table 7: The single parameter EVPPI values for all the parameters in the Chemotherapy model.

3.6.1.1 *A trial to reduce uncertainty in π_1*

In this setting, the trial involves giving the new Chemotherapy drug to 150 patients and measuring the number of patients who experience blood related side effects. For simplicity, we assume complete compliance with the treatment and full follow-up. However, it would be easy to make these assumptions more realistic and generate data with some elements of missingness and then use these datasets to find the posterior. Under these assumptions, the possible future samples are modelled using a binomial distribution conditional on π_1, $X \sim Bin(150, \pi_1)$.

3.6.2 *Determining the optimal number of quadrature points*

In this example, the posterior for the model parameters is found again using MCMC simulation and therefore, to investigate the best value for Q, the total number of posterior simulations was fixed at $500\,000$. This fixes the computation time required to estimate the EVSI across the different values of Q. It also implies that, as Q increases, the number of samples from the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{X})$ decreases. The calculation time for one EVSI estimate is between 17 and 32 seconds on a i7 intel core processor with 32GB of RAM using R $3.2.1$.

Figure 15 presents the sampling distribution of the EVSI estimate with Q between 20 and 100. Again, the EVSI estimate varies because different posterior samples give different values for the posterior variance and therefore varying estimates for the EVSI. Ideally, the sampling distribution of the EVSI would be centered on the true EVSI value, indicated by a black line, with little variance. These sampling distributions represent uncertainty in the variance estimation procedure but exclude uncertainty due to the specific prior sample taken to estimate the prior variance and the sample for moment matching.

Figure 15 demonstrates that a larger Q produces a more accurate EVSI estimate as the distribution is closer to the true EVSI value. However, the computational time does increase slightly are Q increases, even for a constant number of posterior simulations, because the MCMC analysis requires burn-in and initialisation. However, this analysis seems to recommend that Q should be taken larger than 30 to give a more accurate estimate of the EVSI.

At this point, we briefly investigate whether Q should simply be taken as large as possible. In this case, the total number of posterior simulations is reduced to $5\,000$. This restricts the number of posterior simulations used to calculate the posterior variance to between 250 and 50 simulations. Figure 16 demonstrates that the sampling distribution of the EVSI estimates remains constant for $Q > 50$. This implies that increasing Q does not improve accuracy when

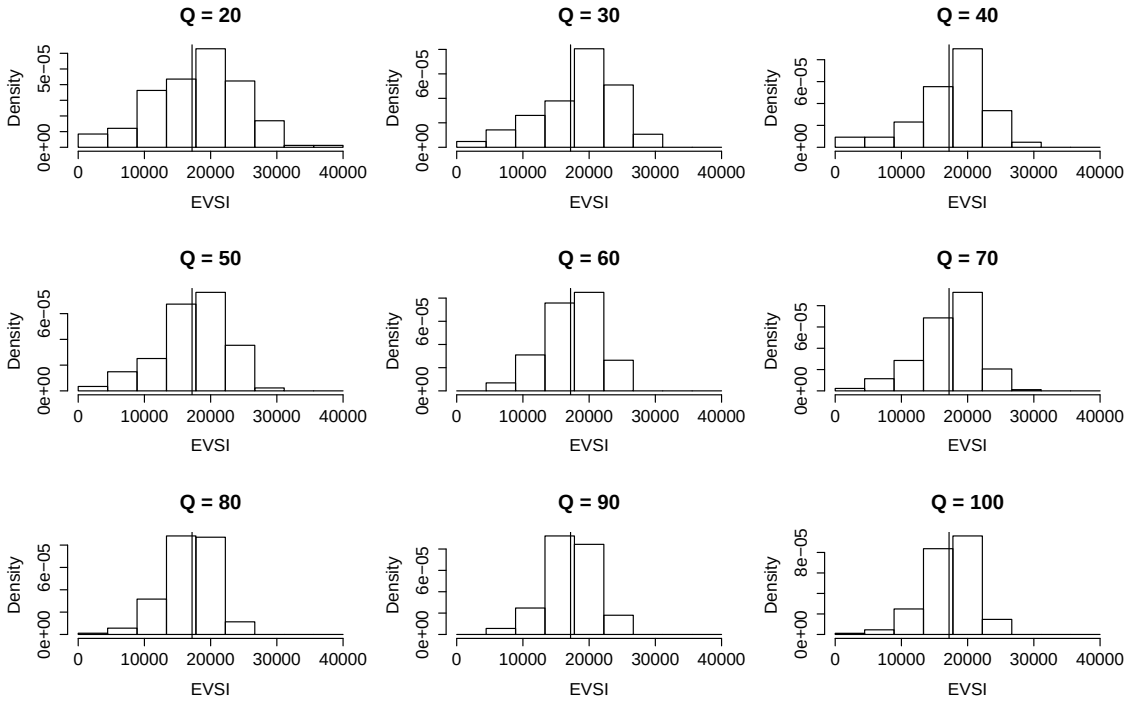


Figure 15: The sampling distribution of the EVSI estimates calculated using variance estimates with different numbers of quadrature points and a total of $500\,000$ posterior simulations. The number of quadrature points for the variance estimate are given in the title. The vertical line represents the MC estimate for the EVSI.

the number of posterior simulations is not sufficient to accurately represent the underlying posterior distribution.

In light of this analysis, Q should be chosen in the following manner. Firstly, determine how many posterior simulations are required to adequately capture the posterior distribution of $\boldsymbol{\theta}$. Then determine, the total computational time available to estimate the EVSI. The value of Q should then be chosen as the total computational time divided by the number of posterior simulations needed to capture the distribution of $\boldsymbol{\theta}$. Provided Q is greater than 30 then this method can be used to estimate the EVSI and will likely provide accurate results, conditional on the sample size of the study being sufficiently large.

3.7 SUMMARY OF THE BASIC MOMENT MATCHING METHOD

Thus far, we have presented an estimation method that reduces the significant computational burden required to estimate the EVSI accurately compared to standard MC simulation. The method is based on the PSA simulations for the model parameters and the net benefit. It estimates the distribution of the preposterior mean using moment matching by performing a linear transformation of the PSA simulations for the incremental net benefit. To perform this matching, nested MC simulation is used to calculate the expected posterior variance across different samples. Nonetheless, this reduces the required number of nested simulations from 10^6 to between $30 - 50$.

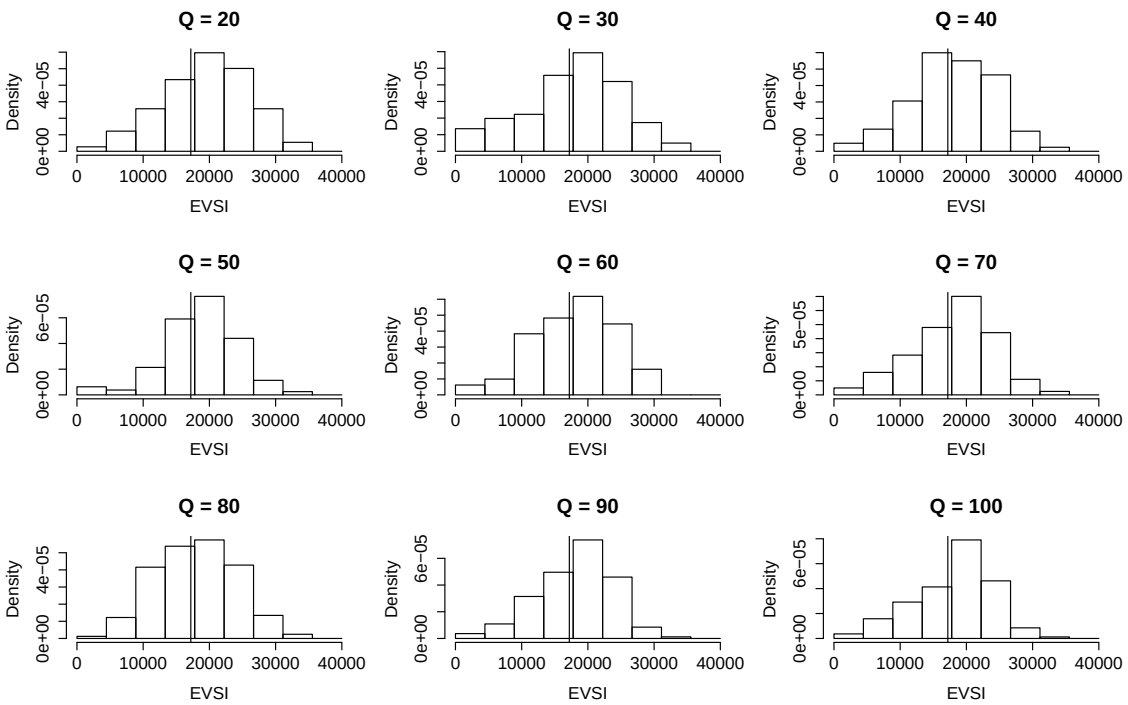


Figure 16: The sampling distribution of the EVSI estimates calculated using variance estimates with different numbers of quadrature points and a total of $5\,000$ posterior simulations. The number of quadrature points for the variance estimate are given in the title. The vertical line represents the MC estimate for the EVSI.

Compared to some of the recent methods for calculating the EVSI (c.f. §3.1), the use of nested simulations implies that this method is relatively slow. However, this additional computational cost allows for a more flexible model specification as it does not require that the data can be summarised in a low-dimensional summary statistic or that the likelihood for the data can be expressed in closed form. It is also easy to consider realistic study designs including missingness, unbalanced designs and complex studies investigating several underlying model parameters.

Perhaps most importantly, the method is based solely on Bayesian updating, which typically must be designed in order to analyse the data once the experiment has been performed. Therefore, calculating the EVSI aids in modelling trial results as researchers can use the procedure that has already been defined. This aligns trial design with trial analysis once the research has been performed.

However, the use of nested simulations makes it challenging to use this methodology for assessing a large number of different trial designs. This is because the nested simulations must be obtained for each of the different designs, vastly increasing the number of simulations needed to calculate the EVSI.

Therefore, we consider an extension to this EVSI calculation method that efficiently estimates the EVSI across different sample sizes. This method is still based on the variance of the posterior distribution but uses Bayesian non-linear regression to avoid sampling Q posterior samples for each sample size and simply requires Q posterior samples in total.

3.8 CALCULATING THE EVSI FOR DIFFERENT SAMPLE SIZES

To extend the EVSI estimation method so it can be used for a range of sample sizes, we begin by considering which sample sizes should be investigated. In general, the trial itself will define practical limits for possible sample sizes, e.g. financial or ethical constraints would give a largest possible sample size, $N_{\max}$, and clinical restriction would imply a smallest possible sample size, $N_{\min}$. Therefore, we consider a vector $\boldsymbol{N} = (N_1, N_2, \ldots, N_Q)$ which contains Q sample sizes between $N_{\max}$ and $N_{\min}$.

As before, we consider generating Q potential future samples from the sampling distribution $p(\boldsymbol{X} \mid \boldsymbol{\theta})$, or more specifically $p(\boldsymbol{X} \mid \boldsymbol{\phi})$, but in this case, the simulation strategy should also change with N, in the simplest setting this would involve simulating N i.i.d samples from $p(X \mid \boldsymbol{\phi})$. To generate these potential samples, we consider the *pair* of values $(\boldsymbol{\phi}_q, N_q)$ and simulate *one* dataset for each pair. This means that we are simulating only one data set for each sample size. These Q datasets are used to update the posterior for the model parameters and calculate the known-distribution net benefit, which in turn is used to calculate the posterior variance of the known-distribution *incremental* net benefit $\text{INB}_t^{\boldsymbol{\theta}}$, denoted σ_q^2.

However, we now use an alternative method to estimate $\sigma_{\boldsymbol{X}}^2$ from σ_q^2. This is because we are now interested in estimating $\sigma_{\boldsymbol{X}}^2(N)$, i.e. the variance of the expected posterior mean $\mu_t^{\boldsymbol{X}}(N)$ for different samples of size N. In a similar manner to Müller and Parmigiani [85], we use regression methods to estimate a function $f(N) = \sigma_{\boldsymbol{X}}^2(N)$, using σ_q^2. The estimated values σ_q^2 are generated conditional on a specific future dataset $\boldsymbol{X}_q$, of size N_q. This dataset represents one possible future, which is subject to variation due to random sampling and dependence on the specific value of $\boldsymbol{\phi}_q$. In the standard setting, we marginalise out this dependence on the specific $\boldsymbol{\phi}_q$ values by taking the mean of σ_q^2. However, in this extended setting, we use regression to marginalise out this dependence across the different sample sizes N_q.

Therefore, we use the values of σ_q^2, obtained by simulation, to estimate the function $f(N)$ using

$$\sigma^2 - \sigma_q^2 = f(N_q) + \varepsilon_q,$$

where ε_q is an error term that captures variation in σ_q^2 due to using simulated datasets with sample size N_q. To present this method, we restrict ourselves to bi-decision settings ($T = 1$) so σ_q^2 is a scalar quantity. In multi-decision settings, σ^2 and σ_q^2 are both matrices and therefore $f(\cdot)$ must be a matrix valued function, complicating the explanation. Therefore, the multi-decision extension of this method is discussed in §4.4.3.

3.8.1 *A Bayesian non-linear model*

To estimate the variance of the preposterior mean for different sample sizes, we suggest a suitable function $f(\cdot)$ and a distribution for the error, along with priors for the model parameters. In the bi-decision setting, we define

$$f(N) = \sigma^2_{\boldsymbol{\phi}} \frac{N}{N+h}, \tag{22}$$

where h is a parameter to be estimated.

The functional form of $f(\cdot)$ reflects information available about the behaviour of the variance of the preposterior mean. Specifically, the variance of the preposterior mean *increases* as the sample sizes increases and $f(\cdot)$ is an increasing function. In addition to this, an infinite sample size for $\boldsymbol{X}$ implies perfect information about the model parameters targeted by the study, and therefore the maximum variance for the preposterior mean is $\sigma^2_{\boldsymbol{\phi}}$. The function $f(\cdot)$, therefore, also has an asymptote at $\sigma^2_{\boldsymbol{\phi}}$. Furthermore, the functional form of $f(\cdot)$ is the exact relationship between the variance of the preposterior mean and N_q in normal-normal conjugate settings with only one underlying model parameter. So, while realistic health-economic models are not based on normal-normal conjugate models, this function is approximately correct in some settings.

To define the full Bayesian model, the residual variance is assumed to be normal; $\varepsilon_q \sim N(0, \sigma^2_{\mathrm{Q}})$ for all $q = 1, \ldots, Q$. Priors are then needed for h and σ^2_{Q}. Throughout this section, we use a non-central half-Cauchy prior for σ_{Q}, as suggested by Gelman [52], where the parameters of the Cauchy-distribution are dependent on the data. This is because the scale of the σ^2_q changes significantly across health economic models. Therefore, the prior mean for σ^2_{Q} is set as half the standard deviation of σ^2_q and the prior variance is the standard deviation of σ^2_q. Finally, a data dependent normal prior, truncated at 0, is set for h with the mean of this normal distribution set to $\frac{N_{\max}}{2}$ and its variance set as $200 \times N_{\max}$, giving vague priors for h and σ^2_{Q}.

3.8.2 *Calculating the EVSI for different sample sizes*

Fitting this non-linear model using the values for σ^2_q requires MCMC and therefore, the posterior distribution for h is estimated by simulation. These posterior simulations can then be used to estimate a distribution for $\sigma^2_{\boldsymbol{X}}(N)$, irrespective of whether N is in N_q, by calculating

$$f(N) = \sigma^2_{\boldsymbol{X}}(N) = \sigma_{\boldsymbol{\phi}} \frac{N}{N+h},$$

for each posterior simulation for h. In theory, these estimates of the variance of the preposterior mean of the net benefit can be used to estimate the EVSI by following steps 9 and 10 in the Moment Matching algorithm in §3.4.4. This would give a posterior sample for the EVSI, where uncertainty in the EVSI reflects uncertainty in the estimation of the variance of the preposterior mean.

However, this process is computationally intensive if we have a large number of posterior samples for σ_N^2. Therefore, it is more computationally efficient to calculate the EVSI for a small number of samples from the posterior of $\sigma_{\boldsymbol{X}}^2(N)$. We advise that the posterior of $\sigma_{\boldsymbol{X}}^2(N)$ is summarised by finding credible intervals, for a small number of confidence levels. The EVSI is then calculated for these posterior quantiles only. For example, in §3.9, the posteriors are summarised by finding the median, the 75% and 95% central credible intervals. Note that the quantiles found for the EVSI using this method are not the posterior quantiles for the EVSI as the relationship between the expected posterior variance and the EVSI is highly non-linear.

Calculating the EVSI using this method only represents uncertainty due to the estimation of h, the regression parameter. Therefore, these distributions exclude all other sources of uncertainty for the EVSI estimation, including estimation based on a specific sample from the prior and estimation based on specific sample from the posterior distribution of the INB. It would be possible to get a model based estimate of the uncertainty in the EVSI estimate due to dependence of the specific samples σ_q^2 by taking the predictive distribution of $\sigma_{\boldsymbol{X}}^2(N)$ directly from the Bayesian model rather than calculating it conditional on the posterior for h. However, this estimate of the error would be incorrect if the regression model for $\sigma_{\boldsymbol{X}}^2(N)$ is poorly specified. This would also ignore error due to the specific prior sample and is unable to estimate the bias due to using the moment matching method to estimate the EVSI.

An important practical consideration for this method is that $\boldsymbol{\phi}_q$ and N_q should be uncorrelated. This is because there is normally a relationship between σ_N^2 and $\boldsymbol{\phi}_q$. In some settings, this can then mask/inflate the relationship between σ_N^2 and N. Thus far, $\boldsymbol{\phi}_q$ has been defined as the sample quantiles for each element of the PSA vector for $\boldsymbol{\phi}$ and N_q as equally spaced values between $N_{\min}$ and $N_{\max}$. However, this specification induces an unintentional correlation between $\boldsymbol{\phi}_q$ and N_q which can bias the results severely. Therefore, it is suggested that each element of $\boldsymbol{\phi}_q$ is randomly reordered to avoid this correlation. Reordering the elements separately also helps to increase the coverage of the $\boldsymbol{X}_q$ values and therefore improve the EVSI estimation.

It is also important to note that this curve fitting method is based on regression. Therefore, it is advisable to check the model fit for the expected posterior variance before calculating the EVSI. In particular, the residuals should be checked for approximate normality and should exhibit no structure, as in standard regression modelling.

3.9 ESTIMATING THIS EVSI USING NON-LINEAR REGRESSION

We implemented this extended method using a hypothetical model [10] and compared to both the standard MC estimator and the EVSI calculation method developed by Menzies [84]. It is then implemented for a practical health-economic example [113] where the EVSI calculations are again compared with the standard MC estimator. Using this final example, we consider the cost of undertaking the trial and use the EVSI to determine the optimal trial design. This analysis is also compared with the standard sample size calculations. Throughout this section, the computational time to complete these analyses is considered.

Parameter	Mean		Standard Deviation	
	$t = 0$	$t = 0$	$t = 0$	$t = 1$
Drug Cost (θ_1, θ_{11})	\$10 000	\$15 000	\$10	\$10
Probability of Hospitalisation (θ_2, θ_{12})	0.1	0.08	0.02	0.02
Days in Hospital (θ_3, θ_{13})	5.2	6.1	1	1
Hospital Cost per Day (θ_4)	\$4 000	\$4 000	\$2 000	\$2 000
Probability of Responding (θ_5, θ_{14})	0.7	0.8	0.1	0.1
Utility Change due to Response (θ_6, θ_{15})	0.3	0.3	0.1	0.05
Duration of Response (years) (θ_7, θ_{16})	3	3	0.5	1
Probability of Side Effects (θ_8, θ_{17})	0.25	0.2	0.1	0.05
Utility Change due to Side Effects (θ_9, θ_{18})	-0.1	-0.1	0.02	0.02
Duration of Side Effects (years) (θ_{10}, θ_{19})	0.5	0.5	0.2	0.2

Table 8: The Parameters for the Brennan and Kharroubi example. The mean and standard deviations for the distributions of the parameters is also given.

3.9.1 *The Brennan and Kharroubi Example*

The first model, developed by Brennan and Kharroubi [10] and modified by Menzies [84], has two treatment options $t = 0$ and $t = 1$ which are both used to treat a hypothetical disease. For each drug, a patient can respond to the treatment, experience side effects or visit hospital for a certain length of time. A utility value is assigned to each of these possible outcomes and costs are associated with the drugs and hospital stays.

All the parameters are assumed to be normal with the mean and standard deviation given in Table 8. In this example, it is assumed that $\theta_5, \theta_7, \theta_{14}$ and θ_{16} are correlated with correlation coefficient 0.6 and the parameters θ_6 and θ_{15} are also correlated with a correlation coefficient 0.6 and independent of the other set of parameters.

The net benefit is given by subtracting the costs from the utility values multiplied by a willingness-to-pay of $k = \$100\,000$. Therefore, the net benefit functions for the two treatments are

$$\text{NB}_0(\boldsymbol{\theta}) = k(\theta_5\theta_6\theta_7 + \theta_8\theta_9\theta_{10}) - (\theta_1 + \theta_2\theta_3\theta_4),$$

$$\text{NB}_1(\boldsymbol{\theta}) = k(\theta_{14}\theta_{15}\theta_{16} + \theta_{17}\theta_{18}\theta_{19}) - (\theta_{11} + \theta_{12}\theta_{13}\theta_4).$$

Five alternative data collection exercises are proposed by Menzies and are also considered in this analysis:

1. A clinical trial collecting information on the probability that a patient responds to the two treatment options which informs parameters θ_5 and θ_{14}.

2. A study looking at the utility improvement for responding to the different treatments which informs parameters θ_6 and θ_{15}.

3. A study investigating the duration of response to the therapy (for those who do respond), informing parameters θ_7 and θ_{16}.

4. A study combining the first two studies, i.e. informing $\theta_5, \theta_6, \theta_{14}$ and θ_{15}.

5. A study combining all the previous studies and therefore informing $\theta_5, \theta_6, \theta_7, \theta_{14}, \theta_{15}$ and θ_{16}.

The likelihood for the data collection is assumed to be independent normal in all settings. The mean of the distribution is then related to the parameter of interest and the standard deviation of the data collection exercise informing each parameter is given in Table 9.

	Data Standard Deviation	
Parameter	$t = 0$	$t = 1$
Probability of Responding (θ_5, θ_{14})	0.2	0.2
Utility Change due to Response (θ_6, θ_{15})	0.2	0.2
Duration of Response (years) (θ_7, θ_{16})	1	2

Table 9: The assumed standard deviation for the data collection exercise targeting uncertainty in the corresponding parameter.

3.9.1.1 *Analysis for the Brennan and Kharroubi example*

To estimate the EVSI for different sample sizes using our method, the known-distribution INB is estimated using one million simulations from the parameter distributions. This implies that the prior variance and mean for INB_1^{θ} is estimated using this full sample. The fitted values conditional on $\boldsymbol{\phi}$, $\text{INB}_1^{\boldsymbol{\phi}}$ were also found using these one million simulations for all the data collection exercises, except exercise 5, using GAM regression §2.3. As exercise 5 has 6 underlying model parameters, GAM regression could not be used and therefore the INLA-SPDE method (c.f. §2.8.1) was used based on $20\,000$ simulations.

In line with Menzies [84], sample sizes between $N_{\min} = 5$ and $N_{\max} = 200$ are considered for each of the different exercises. Throughout the analysis, we set $Q = 50$ which implies that $10\,000$ simulations are taken from 50 different posterior distributions to calculate the variance of the posterior INB_1^{θ} for 50 different sample sizes. The distribution for the EVSI is then determined using the method described in §3.8.2.

The results determined using the EVSI estimation method proposed in this thesis are compared with the conventional approach for calculating the EVSI and the method developed by Menzies. These results are taken directly from Menzies [84] and are the most accurate estimates available. The conventional approach requires one billion model evaluations per sample size compared with $500\,000$ model evaluations for the Moment Matching method to estimate the EVSI across the different sample sizes.

3.9.2 *A Health Economic Model for Chronic Pain*

Sullivan *et al.* [113] developed a cost-effectiveness model for evaluating treatment options for chronic pain. The model has a Markov structure with 10 possible states, where each state has an associated utility score and cost. The disease progression for chronic pain is as follows, the treatment is administered and a patient can either experience adverse effects (AE) or not. The patient can then withdraw from treatment, either due to the AE or otherwise. They can

then continue to another treatment option or withdraw completely from treatment. After the second line of treatment, they either have subsequent treatment or discontinue, these are both assumed to be absorbing states.

In the standard model, used for this example, it is assumed that a patient can either be offered morphine or a new innovative treatment in the first line of treatment. If they withdraw and receive a second treatment, they are only offered oxycodone meaning the only difference between the two treatment arms occurs when the first treatment is administered. The new innovative treatment is more effective and reduces the probability of AE but is more expensive.

The model is built in a UK context and therefore the willingness-to-pay is taken as £$20\,000$. For a more in-depth presentation of all the model parameters, see [113]. The PSA distributions are defined as gamma distributions for costs and beta distributions for probabilities and utilities. The parameters for these distributions are chosen such that the mean of the distribution is informed by a literature review and the standard deviation is taken as 10% of the underlying estimate.

3.9.2.1 *Analysis for the Chronic Pain Example*

A full VoI analysis, as set out by Tuffaha *et al.* [117], is undertaken for the chronic pain model. To begin, $100\,000$ simulations are taken from the prior for the model parameters. These are the fed through the Markov model to give $100\,000$ simulations for $\text{INB}_1^{\boldsymbol{\theta}}$. Calculating the EVPPI based on these simulations indicates that the most valuable parameters are those relating to the utility of the different health states for the first line of treatment. Specifically, the utility of not having any AE from the treatment and withdrawing from the treatment without experiencing AE.

In response to this, we designed an experiment to learn more information about the utility of these two health states, which, if known with certainty, would account for about 79% of the uncertainty in the decision. Questionnaires are a standard method for determining QALYs and therefore, we designed an experiment to send questionnaires to N different participants. We assume, in a simplistic manner, that a participant is interviewed if they withdrew from the first treatment without any AE and they are questioned about both health states which they can accurately recall. This means that the each questionnaire directly updates information about both the key utility parameters.

It is assumed that the two responses to the questionnaire are transformed into QALYs which are then modelled with independent beta distributions that directly informing about the two utilities – not having AE and withdrawing without AE. The means of these beta distributions are set as the utility of interest and the standard deviation of the beta distributions are equal to 0.3 and 0.31 respectively, taken from Ikenberg *et al.* [66]. We consider sample sizes between $N_{\text{min}} = 10$ and $N_{\text{max}} = 150$. It is then assumed that only a proportion of the questionnaires are returned, leading to missingness in the data.

Within this context, two different designs are considered based on an incentive study run alongside the MINT trial [51, 71]; the questionnaire could be sent with £5 incentive to complete. The MINT study demonstrated that sending the incentive improved response rate from 68.7%

to 75.7% while increasing the cost from £4.64 to £9.35. The difference in cost is not equal to the incentive because non-respondents are chased up by telephone which costs staff time and other resources and sending the financial incentive reduces the number of chase-up calls needed.

Therefore, we will use the EVSI to answer two key questions regarding the design of this study; firstly, should the incentive be used? and secondly, how many questionnaires should be sent? The EVSI is estimated for both response rates using $Q = 50$ with 10 000 posterior simulations fed through the Markov model and used to estimate σ_q^2 for each $q = 1, \ldots, Q$. The EVSI results using the moment matching method are compared, in this model, with full MC simulations for sample sizes 10, 25, 50, 100 and 150. These nested MC estimators were calculated using 100 000 PSA simulations to generate the future samples and then 100 000 posterior simulations for each PSA simulation. So, 1 billion simulations were used for each sample size to estimate the EVSI using nested Monte Carlo simulations as opposed to 500 000 simulations for our novel method.

3.10 ESTIMATING THE EVSI ACROSS DIFFERENT SAMPLE SIZES: RESULTS

3.10.1 *The Brennan and Kharroubi example*

Figures 17, 18 and 19 show the EVSI estimates for all the five exercises for the BK example. The solid line gives the EVSI calculated with the median of the posterior distribution of σ_q^2, whereas the dashed line is the 75% credible interval and the dotted line the 95% credible interval. The EVSI estimates from the two-step Monte Carlo estimator and the Menzies estimator are given by the red dots and the blue crosses respectively. The two-step MC estimator (representative of the "true" EVSI) is within the 95% credible interval for all exercises.

Figures 17, 18 and 19 demonstrate that the EVSI is estimated with more relative precision as the EVSI estimate increases. Again, this is because, for small values of the EVSI, the difference between σ_q^2 and the variance of the $\text{INB}_t^{\boldsymbol{\theta}}$ is small so the estimate of the two variances needs to be very accurate to correctly estimate the difference. However, the confidence bands allow researchers to assess the accuracy of the EVSI estimate. If the EVSI estimate is too variable to aid decision making, then more simulations should be undertaken. In general, extra simulations should be gained by increasing Q, provided the number of posterior simulations is sufficient to characterise the distribution of the posterior INB §3.6.2.

3.10.2 *Chronic Pain*

Figure 20 plots the economic value for both studies in the chronic pain example; the gray lines are the economic value of the study with no incentive and the black lines are the value of the incentive study from the moment matching method. In this setting, the economic value is the difference between the EVSI and the cost of the study, from [51]. Clearly, the no incentive study is economically more valuable than the incentive study. This is because the reduced

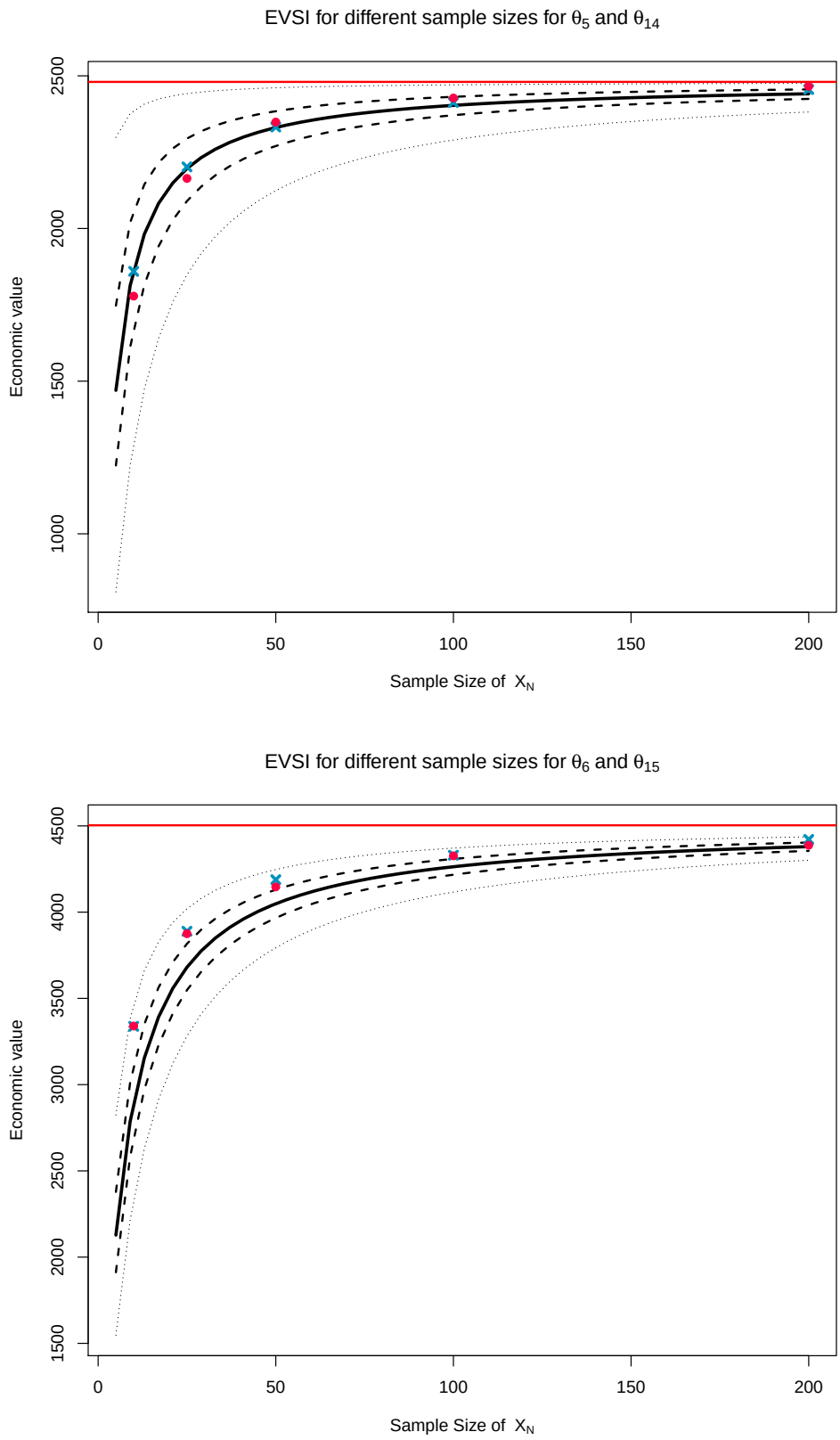


Figure 17: The EVSI estimate with error bounds for data collection exercise 1 (top) and exercise 2 (bottom) in the BK example. These are compared with the 2-level MC estimates (red dots) and the Menzies estimates (blue crosses).

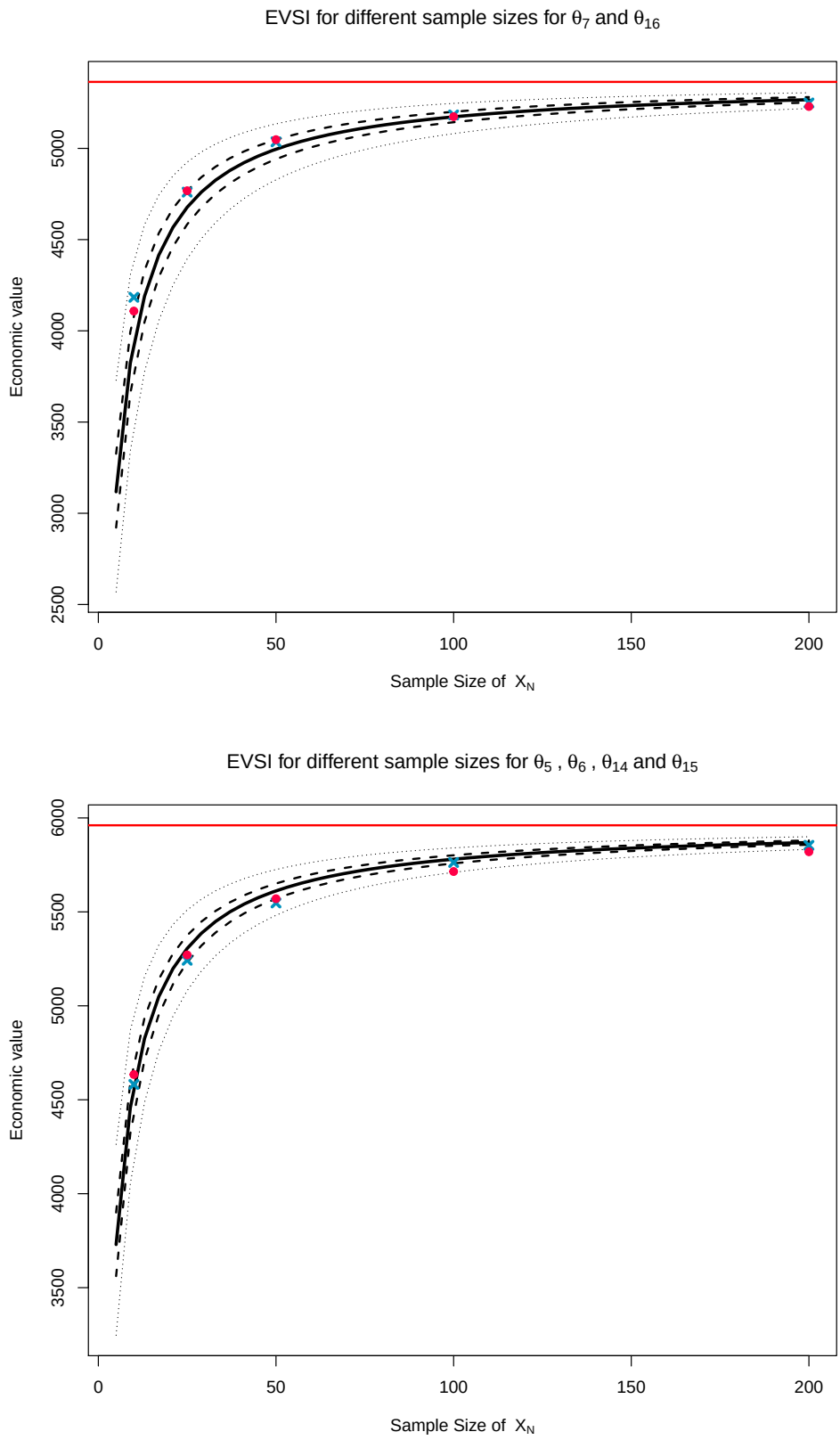


Figure 18: The EVSI estimate with error bounds for data collection exercise 3 (top) and exercise 4 (bottom) in the BK example. These are compared with the 2-level MC estimates (red dots) and the Menzies estimates (blue crosses).

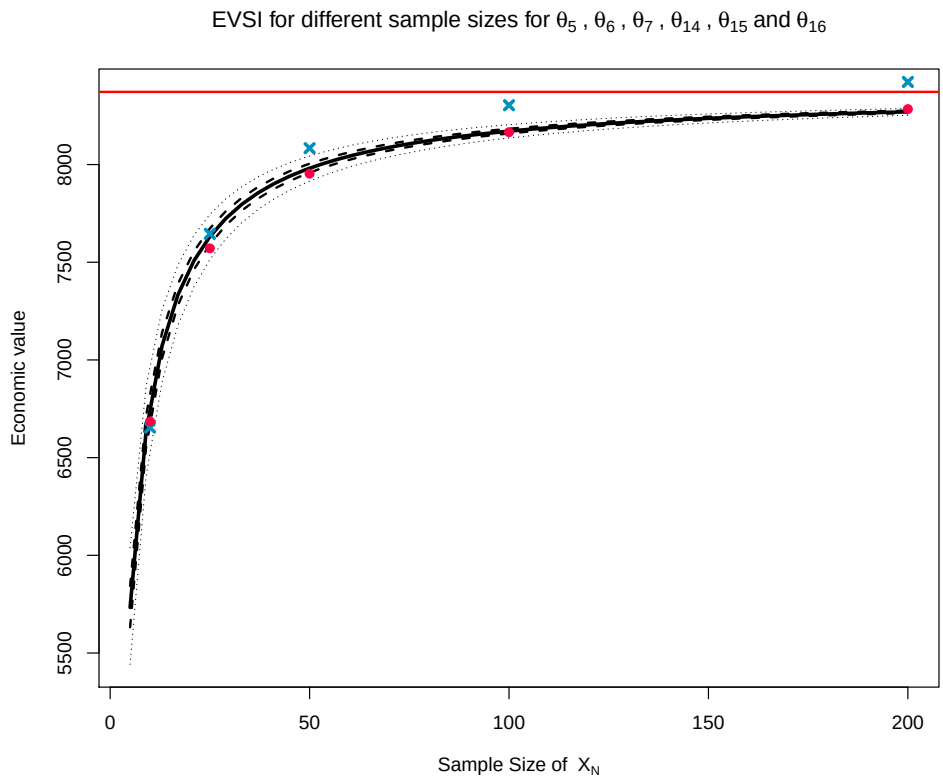


Figure 19: The EVSI estimate with error bounds for data collection exercise 5 in the BK example. These are compared with the 2-level MC estimates (red dots) and the Menzies estimates (blue crosses).

missingness in the questionnaire responses is not economically valuable enough to warrant the cost of the incentive. Note, however, that the dominance of the no incentive study is uncertain for very small sample sizes as the EVSI distributions overlap.

Figure 20 also gives the Monte Carlo estimates for the economic study value; the black crosses give the value for the no incentive study and the grey crosses for the incentive study. Using these estimates we demonstrate that calculating the EVSI using the moment matching method with non-linear regression in line with the nested MC simulation. This demonstrates that, in this example, our novel EVSI calculation method is accurate despite the non-linear and non-Gaussian Markov model structure of the health economic model.

Analysing the economic value for different sample sizes allows us to determine that the optimal sample size in terms of economic benefit is 16 for the incentive study and 27 for the no incentive study. Therefore, in this example, the optimal study is a no incentive study with 27 participants. Once 27 participants have been recruited the additional information gained by sending more questionnaires is no longer worth the additional cost of sending the questionnaire and chasing the participant for a response. In a standard sample size calculation where we aim to half the parameter uncertainty, the required number of *responses* is 75 which is clearly economically inefficient.

The total computational time to calculate the EVSI for both studies using our novel calculation method is 331 seconds or around 5 and a half minutes on a standard desktop computer

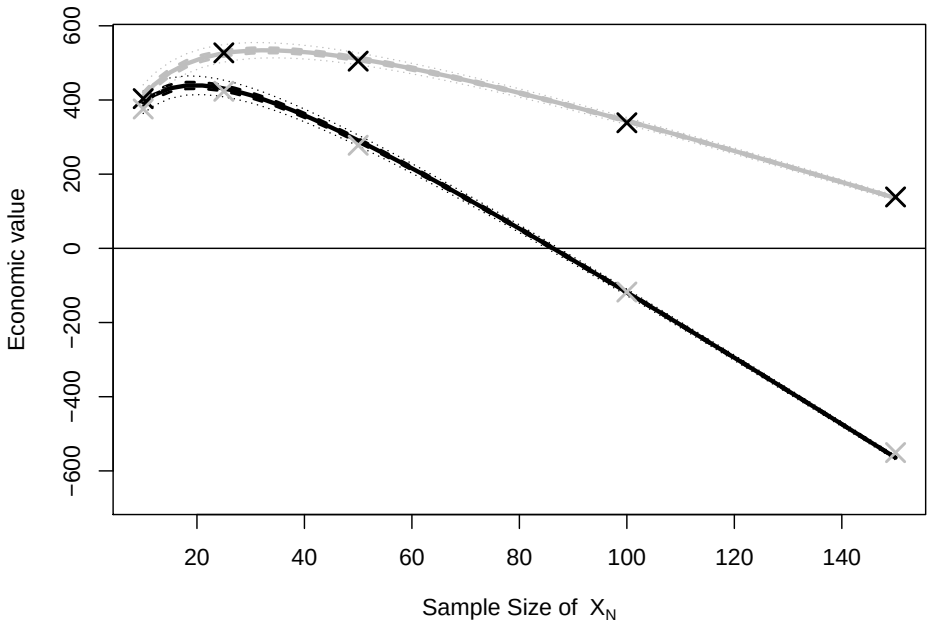


Figure 20: The economic value of the two alternative questionnaire collection strategies for different sample sizes in the chronic pain example. The grey curves indicate the potential value of the study without incentives and the black curves the value of the study with incentives estimated using the moment matching method. The black crosses represent the no-incentive MC simulation estimates and the grey crosses represent the same estimates for the incentive study.

with a Intel i7 Pro processor using R version 3.2.1. This computational time includes fitting the regression model to determine the EVSI by sample size for both study designs, which is around 5 seconds per study. On the other hand, the nested MC simulations required between between 7 and 68 days to calculate each EVSI estimate, with a total computation time of around 311 days for 10 estimates. Therefore, the computational time in this setting is highly competitive compared to other methods even though this study is non-conjugate and the economic model has a Markov structure.

3.11 DISCUSSION OF THE EVSI CALCULATION METHODS

The EVSI has long been touted as an alternative method for performing sample size calculations for designing trials. However, its application has been limited by the immense computational time required to obtain the number of EVSI estimates needed to optimise over different sample sizes. Recently, several approximation methods have been proposed for the EVSI which have reduced the computational time per estimate to allow for this type of analysis. However, the computational time increases linearly with the number of sample sizes required in these standard analysis.

In this chapter, we presented a method to approximate that EVSI based on a small number of nested simulations. The advantage of this method is that it can be extended to allow for the calculation of the EVSI over different sample sizes for a minimal fixed additional computational cost. This means that the cost for a single EVSI estimate and an evaluation for different sample sizes is approximately equal. An additional advantage of this method is that it relatively simple to calculate provided that a model can be defined for the data and that it is possible to sample from the model conditional on the underlying model parameters.

Throughout the chapter, we have applied this method to a large number of health economic models of varying degrees of complexity. We concluded with a full VoI analysis to a real-life health economic model designed to evaluate interventions that treat chronic pain. The analysis determines the optimal study design and sample size required to adequately estimate QALYs relating to treatment for chronic pain. This demonstrates the power of using VoI measures to motivate study design. It also demonstrates that this type of analysis can be achieved within a feasible computational load.

4

PRACTICAL CONSIDERATIONS FOR GENERAL PURPOSE SOFTWARE

The methods presented in this thesis should theoretically allow researchers to use both the EVPPI and EVSI in practice to determine research priorities and optimal trial designs for a large number of health economic models. However, while these methods are valid for calculating these measures, they can be complicated to implement for practitioners working in health economics. Therefore, there is a need for general purpose software to be made available to help with the calculation and presentation of both these VoI measures. This should aid with the comprehension of these measures, which has been cited as another barrier to their widespread use [117].

Therefore, we now detail the technical considerations required to implement these methods in general purpose software. We also discuss some additional methodology to deal with situations where these estimation procedures can struggle, alongside modifications that must be made to allow them to be implemented within `R` functions. We also present some standard graphics for these VoI measures to aid with the presentation to other stakeholders in health economic evaluations.

Finally, practitioners will often wish to use these functions to calculate the EVPPI and EVSI without an in-depth knowledge of the underlying methodologies. This means that we must develop some "debugging" methods for practitioners to determine whether the EVSI and EVPPI have been accurately estimated. Therefore, this chapter also details these debugging methods and how practitioners should improve the approximations when they encounter difficulties.

Some of the description of the `R` package `BCEA` comes from the book entitled *Bayesian Cost-Effectiveness with the R package BCEA* published in the *Use R!* series by *Springer* [5]. Description of the graphics for presenting the EVSI comes from the paper entitled *Development of a New Software Tool to Compute the Expected Value of Sample Information: An application to the HomeHealth intervention* [58] presented at the *HESG Winter Meeting 2018*.

4.1 THE R PACKAGE BCEA

`BCEA` [6] is an `R` [95] package that provides integrated health-economic evaluations based on the PSA samples. The aim of `BCEA` is to produce standardised output, mostly in terms of standard graphics, that can be used to present the results of a health economic evaluation. Therefore, `BCEA` includes functions to calculate both the EVPI and the EVPPI and is also

used to produce graphics for both these measures. In addition to this, BCEA is also coupled with a shiny web application [20] BCEAWeb that allows researchers with no knowledge of R to produce these graphics alongside a standardised report.

4.1.1 *Using BCEA for the EVPI*

Due to the ease with which the EVPI can be calculated by MC simulation, the EVPI was already included in BCEA before the start of this work. Nonetheless, it is briefly presented here for completeness. The main difficulty associated with including the EVPI in the BCEA package is the specification of the net benefit function. Recall from §1.2, that the known-distribution net benefit for each treatment is defined as:

$$\mathrm{NB}_t(\boldsymbol{\theta}) = k\mathrm{E}[e_t \mid \boldsymbol{\theta}] - \mathrm{E}[c_t \mid \boldsymbol{\theta}],$$

where k is the willingness-to-pay that must be specified by the HTA authority. Additionally, we stated that this willingness-to-pay value is normally defined in a range rather than a specific value, for example, NICE suggests taking k between $20\,000$ and $30\,000$ [87] while Claxton *et al.* suggest it should be taken closer to $13\,000$ [26].

The definitions of all VoI measures are dependent on the known-distribution net benefit and therefore dependent on deciding a value for k. Therefore, in BCEA, the EVPI is calculated for a large number of different values of k, defined on a grid. In the standard BCEA procedure, this grid is defined as 501 points between 0 and $50\,000$.

To present these results coherently, the EVPI is plotted across the different values for the willingness-to-pay, as in Figure 21. In general, figures for VoI measures over different

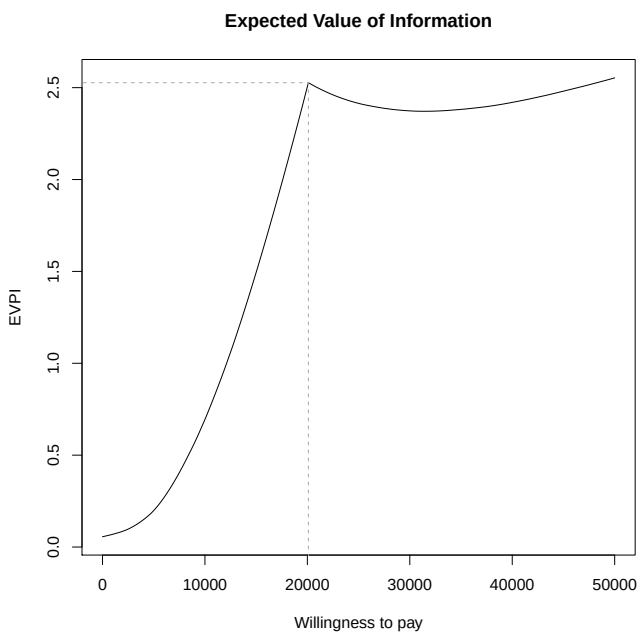


Figure 21: Expected value of perfect information for the Vaccine example §2.4.1.1.

willingness-to-pay values exhibit slightly strange behaviour. Specifically, the graph will normally exhibit break-points, at just over $20\,000$ in Figure 21. This is the point at which the optimal treatment switches. In a large number of health-economic evaluations, the new treatment(s) will be more expensive and more effective than the standard-of-care. If this is the case, then there will be a willingness-to-pay where the payer is then "willing to pay" for the additional level of effectiveness. At this point, the optimal decision switches from the standard of care to the new innovative treatment(s). This creates a discontinuity in the graph for any VoI measure and is associated with the point where the decision uncertainty is maximised.

4.2 CALCULATING THE EVPPI USING BCEA

The SPDE-INLA method for calculating the EVPPI is integrated into `BCEA` in the function `evppi`. Most of the other EVPPI calculation methods, presented in §2, are also integrated into this function, with GAMs the default for single parameters and SPDE-INLA for all larger parameter subsets.

4.2.1 *Calculating the EVPPI for different willingness-to-pay values*

As with the EVPI, it is desirable to calculate the EVPPI for a large number of different willingness-to-pay values. In §2.3, the EVPPI approximation is developed using non-parametric regression to fit a model between the known-distribution net benefit and the parameters of interest. At first glance, this suggests that a standard `BCEA` procedure should calculate the EVPPI using non-parametric regression 501 times, across the different willingness-to-pay thresholds. Clearly, this is infeasible given the computational time of an EVPPI analysis, so this procedure is adapted to integrate EVPPI calculations into `BCEA`.

Specifically, we calculate a separate regression curve for the costs and for the effects,

$$\begin{aligned}\mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}_s}\left[\mathsf{NB}_t(\boldsymbol{\phi}_s,\boldsymbol{\psi})\right] &= \mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[k\,e(\boldsymbol{\phi}_s,\boldsymbol{\psi}) - c(\boldsymbol{\phi}_s,\boldsymbol{\psi})\right]\\ &= k\,\mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}_s}\left[e(\boldsymbol{\phi}_s,\boldsymbol{\psi})\right] - \mathsf{E}_{\boldsymbol{\psi}|\boldsymbol{\phi}}\left[c(\boldsymbol{\phi}_s,\boldsymbol{\psi})\right],\end{aligned}$$

and then combine the fitted values for different thresholds k. This greatly reduces the computational cost of calculating the EVPPI for different values for k as you only have to calculate two regression curves. However, it adds additional complexity compared to §2.3, as we are now estimating two regression curves for each decision.

4.2.2 *Incremental Costs and Effects*

Furthermore, it is possible to reduce the computational complexity of the EVPPI calculation by working with *incremental* costs and effects. Therefore, for each EVPPI calculation with $T+1$ treatment options (where $t=0$ indicates the reference treatment as in §1.1), we would only fit T regression curves. In this setting, the incremental costs and effects for the reference treat-

ment is 0 for all PSA samples. This implies that the conditional expectation for the reference treatment is also 0 irrespective of the parameters of interest $\boldsymbol{\phi}$.

Therefore, in `BCEA`, the EVPPI is found by estimating $2T$ regression curves. The fitted values from these regression curves are combined for different values of k and then the optimal treatment is found by determining which treatment has the largest fitted net benefit. If all the treatments have a negative net benefit, then the reference treatment is optimal. Fitting fewer regression curves improves the numerical stability of this method along with reducing the computational time [111]. As an aside, the computational time results in §3.5.3.2 are the times taken to calculate one regression curve as both examples have two treatment options $T = 1$ and we were working with the incremental net benefit directly.

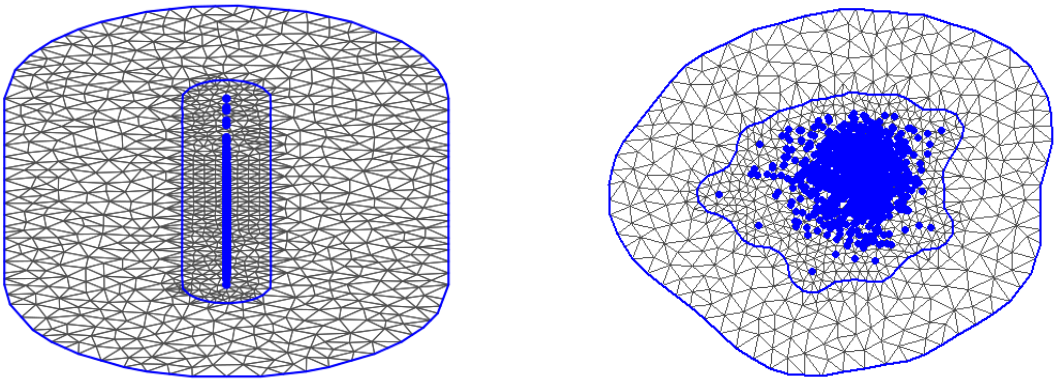
4.2.3 *The SPDE-INLA mesh*

As the SPDE-INLA approximation to the latent Gaussian field is built up on a triangulation (Figure 7), an accurate triangulation must be created. Previously, in §2.8, we discussed the considerations for the triangulation; it must completely surround the data points with a significant distance from the outermost data point to the final boundary, as there are artificial constraints at the boundary of the triangulation. As the computation time for this method is related to the number of grid points, a tight boundary hugs the data points closely and within this boundary the mesh points are dense to give a good approximation. Outside of this inner boundary, the approximation can be rougher and the triangles are therefore larger. The mesh approximation is most efficient when the two dimensions (coming from the projections for EVPPI calculation) are on approximately the same scale. Therefore, the PSA inputs are rescaled in `BCEA` before calculating the projection. This avoids situations such as that shown in Figure 22 (a), where a large number of triangles cover an area with no observations. Rescaling has no effect on the estimated EVPPI value [111] but does significantly decrease the computation time.

The triangulation must be dense enough to adequately capture the underlying structure or the EVPPI estimate will be incorrect. For integration in `BCEA`, we have set a default mesh density that is a twentieth of the range of the projected scaled $\boldsymbol{\phi}$ values. However, larger datasets require denser triangulations to calculate the EVPPI efficiently as typically the number of mesh points should be greater than the number of data points. Therefore, we have allowed the user to change the value for the creation of the mesh.

4.2.4 *The INLA procedure*

As discussed in §2.7, the INLA algorithm explores the hyperparameter space in steps. If these steps are too large then the approximation does not work well and sometimes the numerical results break down. The size of these steps must change dependent on the scale of the underlying response variable and the parameters. In this case, the response variable is the observed incremental costs and effects which, across models, are often on wildly different



(a) (b)

Figure 22: Two grid approximations for the same data set. The LHS shows the triangulation when the variables are left on their original scale, with the projected data points in blue. Notice that there are a large number of triangles in this case, but a relatively small number that surround the data points. In contrast to this, on the right, where the data points are scaled we note that a much larger number of mesh points cover the data, allowing for a more accurate Matérn field approximation for a fixed computational time.

scales. The parameters of interest will also have different ranges. Therefore, it is often impossible to specify one step size taking into account all these different variables. To overcome this issue, the parameters of interest are normalised and the incremental costs and effects are multiplied by a factor such that their range is fixed and small. This rescaling is used so it can be easily reversed once the fitted values are available. This is because the EVPPI must be calculated on the same scale as the original costs and effects.

Within the `evppi` function it is also possible to fine tune the size of steps through the hyper-parameter space. This means it is possible to fix numerical issues that occur when the steps in the parameter space are too small or too large.

4.2.5 *Principal Fitted Components*

In order to determine the PFC, discussed in §2.8.1.1, the best fit inverse regression model must be found as this estimates the sufficient reduction. Additionally, the number of dimensions in the sufficient reduction must be tested as insufficient reductions will likely lead to incorrect EVPPI estimation. Within `BCEA`, the best fit model is found as follows. Firstly, the optimal error structure, amongst the three possible candidates, independent, heteroskedastic or unstructured, is determined. Using a polynomial with degree 2, the three models are fitted and the error structure with the lowest AIC is chosen.

Then the optimal function for the PFC must be found given this error structure. For simplicity, in `BCEA`, we restrict the class of functions by default to polynomials with degrees between 2 and 10. However, ideally, it would also be possible to specify different functions. The best fit function is selected from the 9 possible candidates by choosing the model with the lowest AIC. The net benefit values are centred and rescaled so the range is 2 in order to find the PFC. This is because high degree polynomials create computationally singular matrices for large costs or effects causing the numerical results used to find the PFC to break down.

Using the optimal model, we then find the PFC, and the dimension with the lowest AIC score is then taken as the dimension of the sufficient reduction. If this number exceeds 2 then the user is warned. The function then proceeds to calculate the EVPPI while assuming that the PFC has dimension 2. This gives a slightly different reduction even in the case where 2 dimensions are deemed sufficient. These 2 components are then used to calculate the mesh and the EVPPI.

Operationally, this model selection procedure must be repeated for the costs and effects for each incremental decision option. Therefore, the PFC model fitting procedure must be repeated $2T$ times, which can add to the computational complexity in models with a large number of potential treatments.

4.3 LINEAR PREDICTOR AND RESIDUAL PLOTTING FOR EVPPI ESTIMATION

In practice, the SPDE-INLA estimation procedure for the EVPPI works well in many examples. However, the user may encounter difficulties when the 2-dimensional reduction is not sufficient. If this is the case, the user will receive a warning. However, as the AIC has a tendency to overestimate the size of the sufficient reduction [31] we require a method to determine whether the EVPPI estimation is still suitable.

As we are calculating the EVPPI using regression methods, it is possible to check the fit of the model, and thus the accuracy of the EVPPI estimate, using standard methods for model checking. Therefore, in line with Strong *et al.* [112], we recommend that residual plots are used to determine whether the SPDE-INLA model has captured all the relevant information about the relationship between the costs or effects and the parameters of interest $\boldsymbol{\phi}$.

To demonstrate this analysis, we use the FEAST example from §2.4.1.2. In this case, we are interested in estimating the EVPPI using the parameters of interest (d_2^M, d_3^M, d_4^M, α, p^B, d^S, p^L). Using the Madan *et al.* method presented in §2.1 [77], the EVPPI for these parameters is estimated as 546. However, in this setting, the AIC suggests that the sufficient reduction has greater than two dimensions for some of the incremental costs and effects and the EVPPI is estimated as 523.

Inspecting residuals gives some information about whether the SPDE-INLA model fit is good and therefore whether the function $g_t(\boldsymbol{\phi})$ is estimated correctly. Figure 23 demonstrates clear structure in the residuals, which indicates that the SPDE-INLA procedure is not picking up all the structure in the data. As the SPDE-INLA method is an approximation, it is possible to estimate the EVPPI accurately by reverting to the standard GP method (provided there are

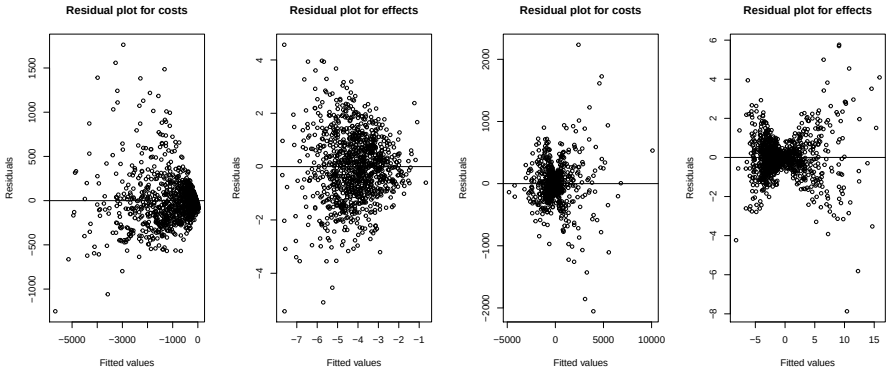


Figure 23: Residual plots for the incremental costs and effects for treatment 3 and 4 in the FEAST example with no interactions in the linear predictor. Treatment 1 is taken as the reference treatment.

some assurances that the hyperparameters have been estimated correctly). Nonetheless, due to the computational complexity of this standard method it is preferable to begin by augmenting the SPDE-INLA method.

In general, the SPDE-INLA fit can be improved by adding additional structure to the linear predictor to correct for some of the information lost through the dimensionality reduction. This means that the $\boldsymbol{H}$ matrix in equation (12) changes its form to include non-linear functions of the parameters. For example, in a three parameter setting the $\boldsymbol{H\beta}$ would have this structure;

$$\beta_0 + \beta_1\phi_s^1 + \beta_2\phi_s^2 + \beta_3\phi_s^3 + \beta_4\phi_s^1\phi_s^2 + \beta_5\phi_s^1\phi_s^3 + \beta_6\phi_s^2\phi_s^3.$$

Allowing for 2nd or 3rd order interactions between the parameters can account for the information lost, whist retaining the computational advantages of the SPDE-INLA method. Evidently, the computational time will increase as the number of hyperparameters to be estimated increases. Specifically, the number of hyperparameters increases with respect to number of parameters in $\boldsymbol{\phi}$, so using interactions when $\boldsymbol{\phi}$ is high-dimensional can be expensive, although still less expensive than the standard GP method.

It may also be possible to use underlying information about the model parameters to reduce the number of interactions needed. For example, there may be evident sub-groups within the parameter set, costs, probabilities, and interactions within these subsets could then be modelled. Most importantly, the use of residual plots allows users to understand whether this additional manipulation is needed.

Introducing 3rd order interactions for each incremental costs and effects measure produces a residual plot (Figure 24) that is slightly closer to the expected structure for some of the incremental costs and effects, and gives an EVPPI estimate of 543 which is very close to the true EVPPI value. However, there are still some clear issues with the fitted values for the incremental effects for the 4th treatment. The same behaviour is also exhibited for the incremental effects for treatment 2 (Figure not shown). The fitted values for the incremental costs for treatment 3 also exhibit some strange structure for small values of the fitted costs.

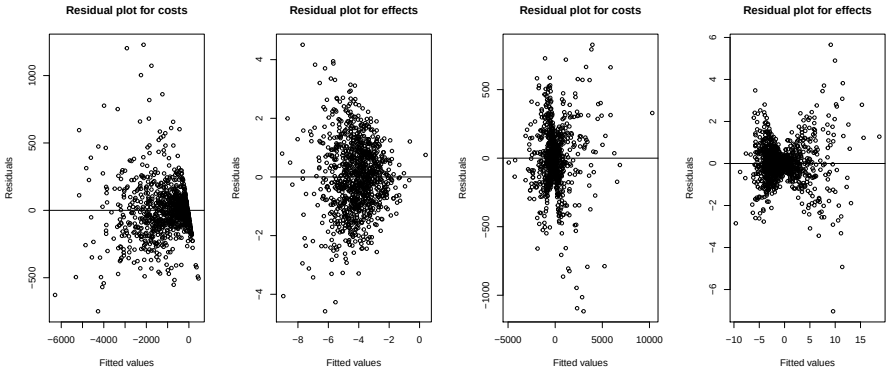


Figure 24: Residual plots for the incremental costs and effects for treatment 3 and 4 in the FEAST example with 3rd order interactions in the linear predictor. Treatment 1 is taken as the reference treatment.

Therefore, we advise that the standard GP methodology is used for these three regression fits, while we retain the SPDE-INLA method to find the other three regression curves. This gives an EVPPI estimate of 551 – in this case, surprisingly, this estimate is actually a little further from the truth, despite better model fit.

In general, this is an example of "mixed strategy" EVPPI estimation and this is our recommendation when the sufficient reduction has greater than 2 dimensions. This allows researchers to retain the computational efficiencies of the SPDE-INLA method in situations where it is valid but resort to more accurate and computationally costly estimation methods when needed. We suggest that residual plots should be used to determine whether the estimation is suitable.

Once the EVPPI has been estimated then `BCEA` includes a plot function for an `evppi` object (Figure 25) which displays both the EVPPI and the EVPI. This graphic is useful as the EVPPI is typically used for ranking different parameter sets in terms of importance. In this sense, it is useful to index the EVPPI against the EVPI to determine whether the parameter subset is a large contributor to decision uncertainty.

4.4 SOFTWARE TO CALCULATE THE EVSI

In general, the theory and calculation of the EVPPI is relatively simple compared to the EVSI. Specifically, as the EVSI is concerned with designing a specific experiment it is more challenging to create generally applicable software. This is because the user must be able to design their experiment rather than basing the analysis solely on the PSA samples which will already be available. In addition to this, the number of potential design considerations is very large. For example, we have seen that researchers vary the willingness-to-pay in VoI calculations. However, for the EVSI we can also think of varying the number of patients in the future trial and other considerations such as the number of people who will benefit from the treatment or the length of time the treatment will be available.

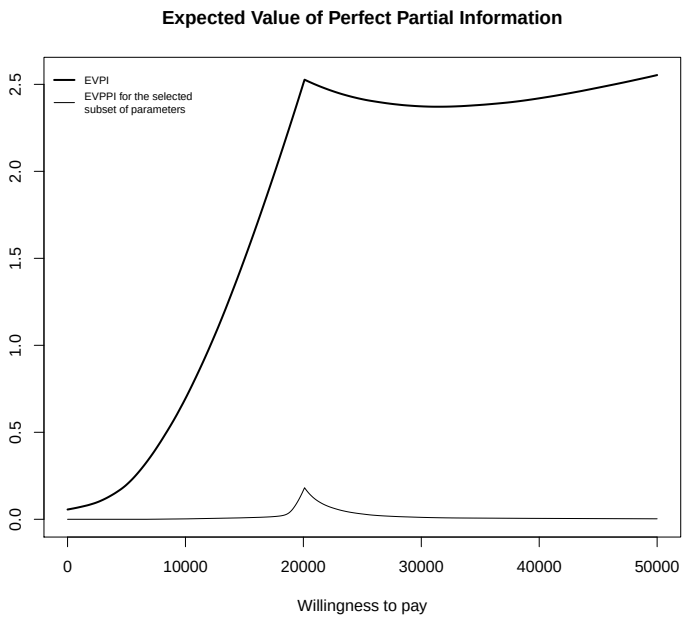


Figure 25: An example of the graph produced when an `evppi` object is plotted in `R`.

Therefore, we have created a new `R` package known as `EVSI` [56] which includes the moment matching method for calculating the EVSI along with a suite of graphics to explore different trial design considerations. The following section begins by discussing the functions for calculating the EVSI. This includes modifications to allow the EVSI calculation across different willingness-to-pay values and extensions to multi-decision settings. We then discuss the different graphics that can be used to present the EVSI, taking into account the different design issues.

4.4.1 *Calculating the EVSI for different willingness-to-pay values*

As each EVSI calculation is based on $Q > 30$ posterior updates, the calculation for different willingness-to-pay thresholds must avoid rerunning the full calculation method for each threshold. Again, this can be achieved by exploiting the linear definition of the known-distribution net benefit;

$$\mathsf{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[\mathsf{NB}_t(\boldsymbol{\theta})\right] = k\,\mathsf{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[e_t(\boldsymbol{\theta})\right] - \mathsf{E}_{\boldsymbol{\theta}|\boldsymbol{X}}\left[c_t(\boldsymbol{\theta})\right].$$

In this setting, moment matching is used to estimate samples from the distribution of the incremental costs and effects separately. The resulting samples are then combined for each threshold and used to calculate the EVSI. This adds little complexity to the methods as they have already been presented. Note, as in all other settings, we work with the incremental costs and effects throughout the EVSI estimation procedures.

4.4.2 *Sampling from the prior-predictive distribution*

To estimate the EVSI, it is necessary to simulate potential data sets $\boldsymbol{X}_q$. In Chapter 3, the potential future datasets $\boldsymbol{X}_q$ for $q = 1, \ldots, Q$ are simulated by taking the quantiles of $\boldsymbol{\phi}_q$ and generating one sample from $p(\boldsymbol{X} \mid \boldsymbol{\phi}_q)$. To use this simulation method within a function, the user must specify a sampling distribution for the data and a model to update the parameters.

However, as the sampling distribution of the data is defined in the Bayesian model for updating the parameters, it is possible to simulate the future datasets directly from the prior-predictive distribution for $\boldsymbol{X}$. Therefore, in this software, users are simply required to supply a Bayesian model to determine the posterior of the parameters, along with functions that calculate the costs and effects, conditional on those parameters. The function then begins by running the Bayesian model with unknown data $\boldsymbol{X}$. This allows us to directly simulate from the prior-predictive distribution of $\boldsymbol{X}$ and then use these simulated samples to update the posterior of the model parameters.

The current method for selecting samples from the prior-predictive distribution of $\boldsymbol{X}$ proceeds by randomly selecting a parameter in the $\boldsymbol{\phi}$ vector for each $q = 1, \ldots, Q$. The $\frac{q}{Q+1}^{\text{th}}$ quantile is then found for the selected element of the $\boldsymbol{\phi}$ vector. If we consider a matrix containing the MCMC samples for both the model parameters and the data, then we select the sample from the prior-predictive distribution that is found in the same *row* as the quantile for the selected element of the $\boldsymbol{\phi}$ vector. In this sense, we are selecting the prior-predictive sample that is generated when the selected element of $\boldsymbol{\phi}$ is equal to its quantile value. This is not an ideal method of selecting elements from the prior-predictive sample of $\boldsymbol{X}$ and future work will focus on more suitable methods to find these samples. One suggestion is to use clustering to span the prior-predictive space efficiently.

4.4.2.1 *Prior-predictive sampling across multiple sample sizes*

When we estimate the EVSI across different sample sizes, the prior predictive distribution must be found for each sample size. This can add to the computation time as the Bayesian model must be run $2Q$ times to achieve the EVSI estimate, first to find the prior-predictive distribution of $\boldsymbol{X}_q$ and then to update the posterior conditional on this dataset. Therefore, for computational reasons, the user can supply the future datasets, rather than sampling them directly within the function. This is because also becomes significantly more challenging to determine what is meant by "spanning" the prior-predictive sample space in this case. Currently, a similar method to the standard setting is used to determine the prior-predictive samples. The main difference is that the order in which the different quantiles are found are randomly reordered to avoid a relationship with N. In practice, the optimal selection method in this setting may be random and therefore future work should focus on the ideal method for selecting these prior-predictive samples within the `EVSI` package.

4.4.3 *Calculating the EVSI in multi-decision problems*

Estimating the EVSI in multi-decision health economic models (i.e. when more than 2 treatments are under consideration) using the simple moment-matching matching is very similar in the dual-decision setting. The only difference is that the variance of the known-distribution incremental net benefit is a matrix in the multi-decision setting. Therefore, the EVSI is calculated using the inverse and inverse square root of the variance matrix, however, as these are well-defined, they can be easily used.

In practice, working with variance matrices (as opposed to scalar values) adds a little computational complexity. This is because the moment matching method is based on calculating the difference between two variance matrices, both calculated by simulation. Occasionally, this leads to a variance matrix for the preposterior mean costs or effects that is not positive definite. In these settings, it is necessary to determine the "nearest" positive definite matrix to the matrix that has been estimated by simulation. This is found using the eigen-decomposition of the estimated "variance" matrix V. Specifically, the nearest positive definite matrix V^* is created by calculating

$$V^* = MD^*M^T,$$

where M is the matrix of eigenvectors of V and D^* is a diagonal matrix of the eigenvalues of V where any negative eigenvalues are set to 0 [63]. This eigen-decomposition is also used to find the matrix inverse and square root of V^* in a computationally efficient manner.

Calculating the EVSI across sample sizes in a multi-decision problem is more problematic. The non-linear model defined in equation (22) is a scalar function and the variance of the preposterior mean is a matrix in multi-decision settings. Therefore, we must extend this regression model to a matrix valued function. In practice, we extend the non-linear regression function by fitting the non-linear model from equation (22)

$$f(N) = \frac{N}{N+h}$$

separately for each unique element of the variance matrix. Again, it is possible to demonstrate in a decision model with three treatment options that this is approximately accurate in normal-normal conjugate settings. Specifically, the constant h is slightly dependent on N through the expression:

$$h = \frac{aN+b}{cN+d},$$

where a, b, c and d are all constants that depend on the prior and posterior variance of the costs or effects.

Using this specification for the function $f(\cdot)$, we fit

$$\frac{T(T+1)}{2} \times 2 = T(T+1)$$

regression models separately to calculate the EVSI across different sample sizes, where T is the number of incremental decisions. These curves are then used to calculate the variance-

covariance matrix for the costs and effects. Again, this allows us to calculate a and b from equation (20) and therefore find the distribution of the preposterior mean for the incremental effects and costs. Finally, these are used to calculate the EVSI, following the algorithm in §3.4.4.

As an additional complication, recall that the distribution of the EVSI estimate was approximated by a low-dimensional summary of the distribution of the posterior distribution for h. This method becomes more complicated when multiple curves are being fitted. Nonetheless, in the `EVSI` package, we have simply summarised these distributions by finding the quantiles for each curve and then used these to create the variance-covariance matrices for the costs and effects. These are then used to find possible distributions for the preposterior mean of the costs and effects and then the EVSI is calculated from these distributions. This method clearly does not give credible intervals for the EVSI but simply a low-dimensional summary of the possible EVSI values.

4.5 VISUALISATIONS FOR THE EVSI

Visualisations for the EVSI are more complicated than for both the EVPI and EVPPI as there are more variants in an EVSI analysis. Therefore, a suite of graphics has been included in the `R` package `EVSI` to aid researchers to explore different aspects of the EVSI analysis. Most importantly, a `shiny` web application [20] has also been included in the `EVSI` package. Again, this web interface allows researchers with no knowledge of `R` to explore the graphics presented below and can either be launched from within `R` or is available at `https://egon.stats.ucl.ac.uk/projects/EVSI/`.

Firstly, a graphic is available to plot the EVSI across willingness-to-pay, in a similar manner to the graphics seen for the EVPI and EVPPI, displayed in Figure 26 (a). This compares the relative sizes of the EVSI, EVPPI and EVPI and demonstrates whether the sampling strategy efficiently updates information about the model parameters. Evidently, this graphic is more complicated for the EVSI as it will likely have been computed across different sample sizes. This means that it is possible to plot this graphic for every sample size which can result in a large number of plots.

Therefore, it is possible to produce a graphic where different sample sizes can be displayed at the same time, Figure 26 (b). This allows the user to ascertain how quickly the EVSI approaches the asymptote of the EVPPI. However, as this graphic can be difficult to read, especially when a large number of alternative sample sizes are considered, we allow the sample size to be changed dynamically in the `shiny` web application so the researcher can efficiently explore the different possible sample sizes for the potential future study.

4.5.1 *Displaying uncertainty in the EVSI estimate*

As discussed in §3.8.2, if the EVSI is estimated across different sample sizes using non-linear regression, then the posterior of the h parameter is used to generate a low-dimensional sum-

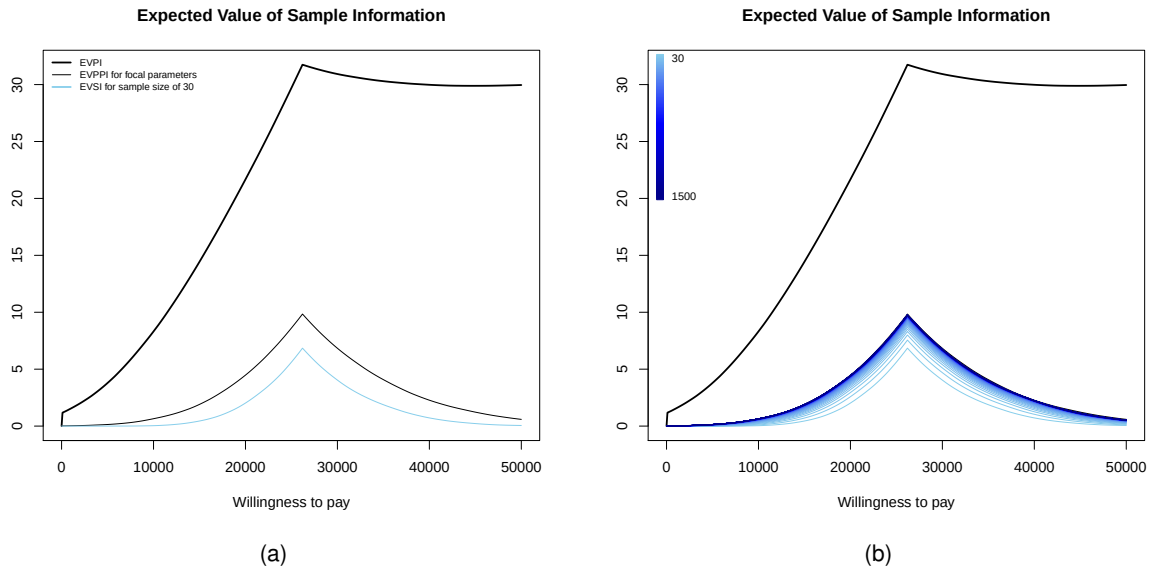


Figure 26: The EVSI plotted across different willingness-to-pay thresholds, (a) shows the EVSI for one sample size and (b) across a large number of sample sizes.

mary of EVSI values. In some sense, this summary represents uncertainty in the EVSI that arises from the estimation procedure. In many settings, it is useful to explore this uncertainty graphically. It is also beneficial to explore the behaviour of the EVSI estimates across different sample sizes. In this setting, both the EVSI estimate and the uncertainty bands will change as a function of the willingness-to-pay threshold.

Figure 27 shows the EVSI estimate across different sample sizes for a fixed willingness-to-pay threshold, along with uncertainty bands. The EVSI will always be increasing and bounded above by the EVPPI and we can clearly see this behaviour on the graphic. A key feature of the moment matching method is that the EVSI becomes relatively more uncertain as the EVSI decreases, which can also be explored using this graphic especially using the `shiny` web application which allows users to change the willingness-to-pay threshold dynamically. In general, decisions regarding future funding decisions should not be made based on the EVSI estimate if the uncertainty bands are wide. In this case, Figure 27 displays an EVSI estimate with relatively tight uncertainty bands.

If the EVSI has been calculated using an alternative calculation method, then it can still be displayed across different sample sizes using this graphic. In this setting, the graphic would not give information about the uncertainty in the estimation procedure but simply allow researchers to determine what sample sizes give close to perfect information. This graphic also checks that the EVSI remains below the EVPPI for all sample sizes which gives some assurance that the estimation procedure is accurate.

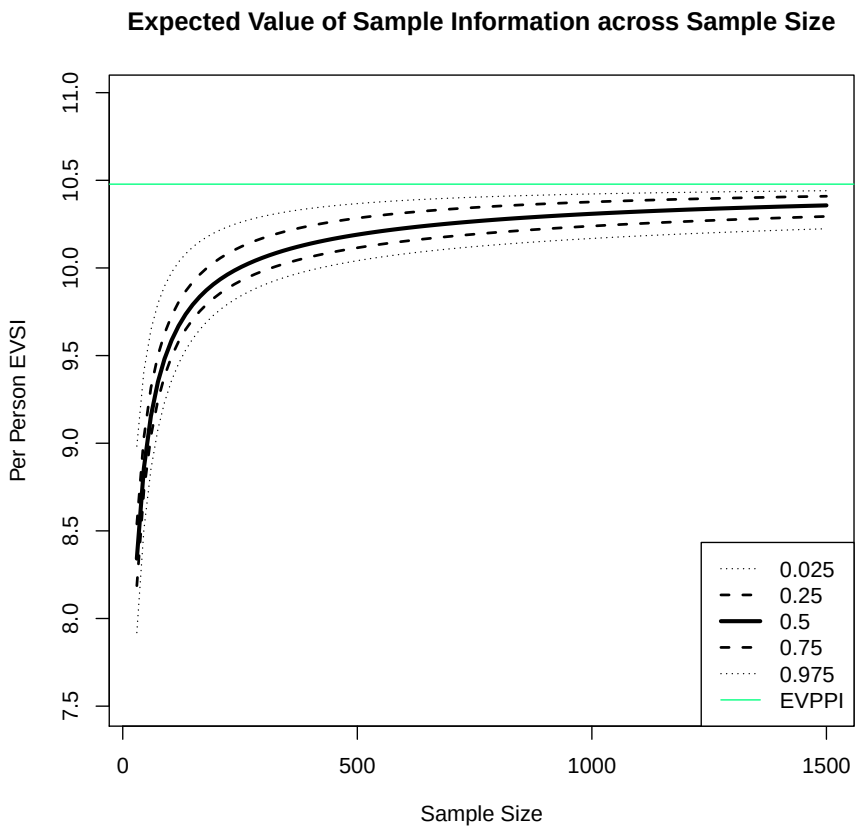


Figure 27: The EVSI plotted across different values of N for a specified willingness-to-pay value. The graphic also shows uncertainty bands for the EVSI.

4.5.2 *Expected Net Benefit of Sampling*

The Expected Net Benefit of Sampling (ENBS) is the difference between the EVSI and the cost of undertaking the study. This gives the net economic *value* of undertaking the study in question. In the first instance, we consider whether the ENBS is greater than 0, i.e. will the proposed study be of value to society/the payer? If the ENBS is positive, then researchers are often interested in finding the optimum study design. There are many different possible designs that could be evaluated but as it is cheap to estimate the EVSI across different sample sizes using the moment matching method, we include functions to find the optimal sample size for the proposed study.

Normally, the EVSI is calculated from an individual perspective, in fact, the y-axis on Figure 27 has the label "per person EVSI". This means that the known-distribution net benefit is calculated as the net benefit of the treatment for each individual who would be affected by the disease. So, the EVSI is the value of sample information to each person that would be affected by the disease. However, if the proposed study is run, the information gleaned from that study will impact the entire population of people with the disease, known as the incidence population. Therefore, to calculate the ENBS, we must multiply by the number of people who will benefit from the treatments under consideration [115].

In practice, multiplying the EVSI by the incidence population, implies that payers should fund trials in high prevalence diseases more frequently then those targeting low prevalence diseases. This is because VoI measures are concerned with the risk of implementing an inefficient treatment that appears efficient under current information. If an inefficient treatment is only going to be used by a small number of patients then this waste of resources may not be severe even if the treatment turns out to be economically inefficient. Therefore, the more widespread the disease, the more people will use the treatment and the more money will be wasted if the treatment turns out to be inefficient. Therefore, more information is needed before making a decision about which treatment to fund in high prevalence settings.

In addition to multiplying by the incidence population, we must also consider that the treatment will be used for several years. In line with standard economic analysis, discounting should be used to imply that immediate benefits of the treatment are more valuable than future benefits [87]. This discounting is applied up to a time at which it is assumed that a new treatment would enter the market and dominate the treatments under consideration. This is known as the time horizon of the treatment [82] and changes from intervention to intervention. In some disease areas, such as oncology, the time horizon is short, while other disease areas are less fast moving and treatments will be used for a longer period of time.

In the `EVSI` package, continuous discounting is used to estimate the ENBS, which is calculated as

$$\frac{P}{d}\left(1 - e^{-dT}\right) \text{EVSI} - C_{\text{study}}$$

where P is the incidence population, d is the discount factor (normally, 3.5% [87]), T is the time horizon of the treatment and C_{study} is the cost of undertaking the study. Evidently, this definition is heavily dependent on the incidence population, time horizon and study cost, where all of these values are subject to uncertainty. For example, the number of years before a competitor treatment reaches the market is rarely known with certainty, neither is the exact number of people affected by a disease.

4.5.2.1 *Probability of a cost-effective trial*

Ideally, the ENBS would be used to calculate whether a trial should go ahead. However, as there is an element of uncertainty in all the key elements of the ENBS definition (including uncertainty in the EVSI itself), we have developed a graphic to explore the uncertainty in all these different elements. Specifically, we use a graphic with incidence population and time horizon as the two axes which then displays a heat-map for the probability of a cost-effective trial (Figure 28), where uncertainty in the ENBS is due to uncertainty in both the EVSI and the study costs.

Practically, the user specifies maximum and minimum possible values for the trial costs, these are usually available as they must be included in applications for trial funding. More specifically, if the EVSI has been estimated across different values for N, users specify the set up costs for the trial, typically sunk costs such as equipment or training that will be spent no matter the size of the trial, and per person costs, costs relating to recruitment and retention

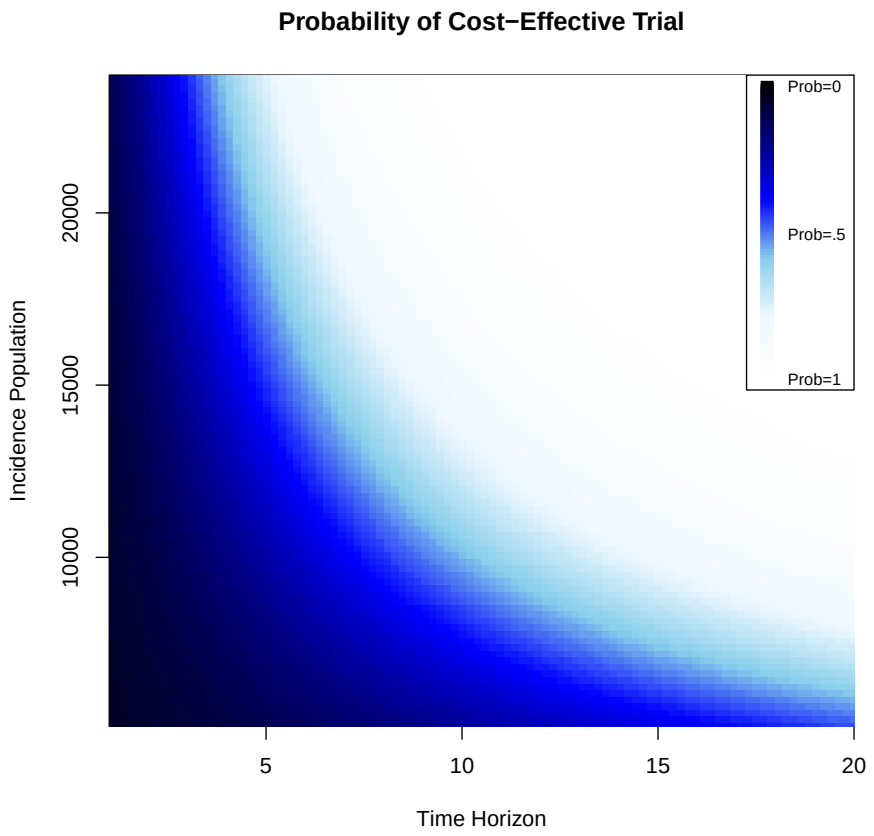


Figure 28: The probability of a cost-effective trial for different lengths for the time horizon and levels of incidence population. The probability represents the uncertainty in the study costs and EVSI estimates.

of each study participant. This allows us to calculate the ENBS across the different sample sizes. It is then assumed that the expected trial costs are normally distributed, centred on the midpoint between the maximum and minimum trial costs. The standard deviation of this distribution is computed by assuming that there is a 95% chance that the trial costs fall within the maximum and minimum. Finally, the EVSI distribution is also approximated with a normal distribution so the distributions for the EVSI and the costs can be combined analytically.

Figure 28 then allows the user to find the probability of a cost-effective trial for different combinations of incidence population and time horizon. For example, the study is unlikely to be cost-effective if the expected time horizon is less than 5 years, but if the time horizon is nearer 20 years then the study can go ahead almost irrespective of the number of people affected by the disease.

Figure 28 is generated for a specific study size and a specific willingness-to-pay value, both of which are also unknown. Therefore, this graphic is included in the `shiny` web application. Again, users can easily change both the sample size and the willingness-to-pay dynamically to explore the probability of cost effectiveness for a large number of alternative specifications of the underlying health economic model and study design. Note that if both the EVSI and

the study costs are known with certainty, Figure 28 simply gives a stop/go decision for each combination of time horizon and incidence population.

4.5.2.2 *Optimal study design*

Finally, it is possible to find the optimal sample size of a proposed study. Again, this analysis is subject to the same uncertainties seen previously, namely uncertainty in willingness-to-pay, incidence population, time horizon, cost of the study and the EVSI. Nonetheless, for a fixed incidence population, time horizon and willingness-to-pay, it is possible to find the optimal sample size and produce a graphic that represents the uncertainty in the ENBS estimate. This translates into uncertainty about the optimal sample size.

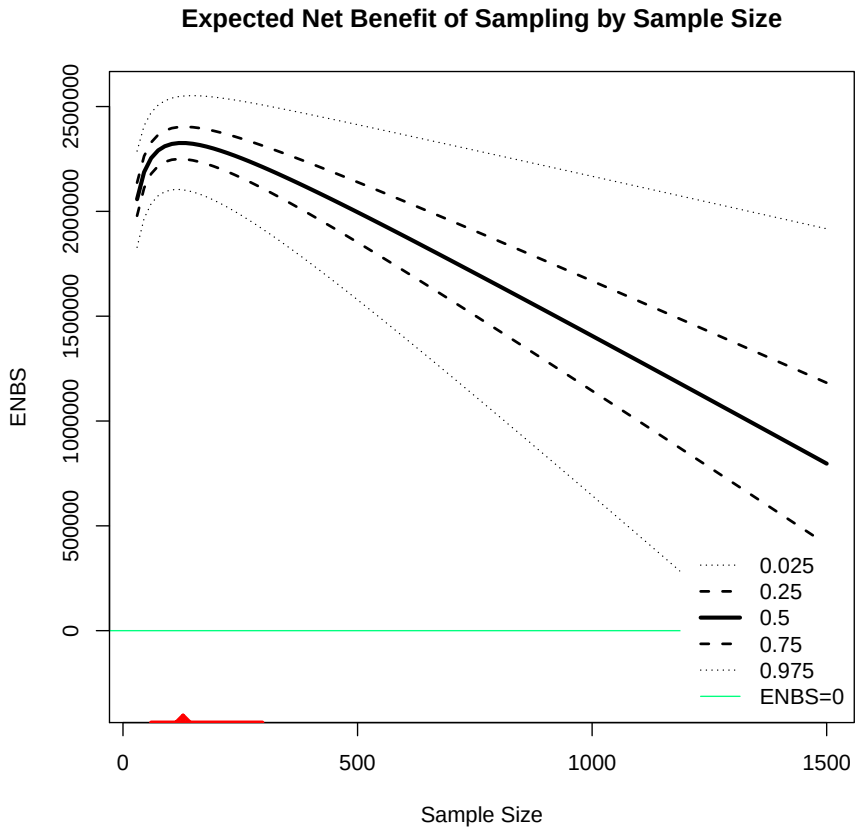


Figure 29: The ENBS across different sample sizes. The alternative curves represent the uncertainty in the value of the ENBS arising from uncertainty in knowledge about the study costs and the EVSI estimation procedure.

Figure 29 demonstrates the proposed graphic for this optimal sample size analysis. Firstly, the ENBS rises to an optimal value before descending when the cost of enrolling additional patients in the trial exceeds the additional value given by the patient. In Figure 29, the ENBS is always positive, so even though an optimal sample size exists, a trial of any of the considered sizes would have an economic benefit, for this willingness-to-pay, incidence population and time horizon. The red triangle on the x-axis marks the sample size at which the ENBS reaches its maximum, 128 patients in this example. The red line then highlights the sample sizes for

which the ENBS is within 5% its maximum. In most settings, the ENBS is a relatively "flat" function, i.e. there are a large number of sample sizes for which the cost of an additional patient is approximately equal to the extra value that a patient would bring. Therefore, we suggest that any sample size within this red region (between 60 and 297) would be suitable.

In practice, the ENBS can therefore be used alongside standard sample size calculations. These determine a trial size by considering how many patients would be needed to find a statistically significant result in the key clinical outcome. It is likely that users will perform these sample size calculations before/alongside the EVSI and then wish to confirm whether this sample size is within the "red" region.

Again, the web application displays this graph dynamically for different combinations of willingness-to-pay values, incidence populations and time horizons. The web application also outputs the numerical values for the optimal sample size and the 5% interval for the sample size. Thus, this graphic can be compared with these numerical values and the uncertainty bands can be used to visually inspect both the probability of cost-effectiveness and the certainty surrounding this optimal sample size.

4.6 CONCLUSION

This chapter has concluded the work on VoI measures by presenting two software packages that have been designed to implement the methods developed in this thesis. Both methods required slight modifications before they can be integrated into software, particularly related to estimating VoI measures across different willingness-to-pay thresholds. This software allows researchers without an in-depth knowledge of the VoI estimation methods to use VoI to inform research and trial design decisions.

Another key element of the software is the inclusion of standard graphics for the presentation and interpretation of VoI measures. Standardised presentation is especially important for the EVSI and ENBS analysis as this often requires researchers to assess these measures across a large number of model and trial specifications. The final contribution to the presentation of these measures is the development of two `shiny` web applications that allow the exploration of standard graphics without any knowledge of `R`.

5

SUMMARY AND CONCLUDING REMARKS

This thesis presented novel methodology to estimate VoI measures using simulation. Specifically, we have focussed on the EVPPI, which gives the economic value of gaining perfect information about a subset of the heath-economic model parameters, and the EVSI, which gives the economic value of a specific study designed to reduce model uncertainty. Both these measures could be used as a tool for research prioritisation and the EVSI can be extended to determine "optimal" trial designs from a health economic perspective.

To develop these new methods, Chapter 2 began by investigating the currently available methods for calculating the EVPPI. This investigation concluded that estimating the EVPPI using non-parametric regression, first introduced by Strong *et al.* [111], is accurate and computationally efficient when the number of parameters of interest is small (< 4). It also concluded that an alternative method should be developed for higher dimensional settings, especially when a large PSA dataset is available.

Therefore, in §2.6, this thesis developed a fast method for Gaussian Process regression which exploited results from the spatial statistics literature, the INLA algorithm and dimension reduction to calculate the EVPPI efficiently. In Chapter 4, we then explored some extensions that allow us to approximate the EVPPI efficiently even when this SPDE-INLA method breaks down. These extensions use standard model fitting techniques to determine whether the fast Gaussian Process regression is still sufficiently accurate and then suggests that a combination of different regression methods should be used to calculate the EVPPI.

Based on this work, we developed the `evppi` function within the `BCEA` package so that the EVPPI can be estimated within a feasible time frame for almost all health-economic models, provided PSA simulations are available. Importantly, the `evppi` function only requires a small number of user inputs, so practitioners can use this function to calculate the EVPPI in practice, without a full understanding of the SPDE-INLA method. The `BCEA` package also couples EVPPI calculations with standard graphics to help key stakeholders understand and digest VoI analysis.

To further extend the use of VoI analysis, Chapter 3 develops a novel algorithm for calculating the EVSI. In basic settings, the posterior expectation of the known-distribution net benefit across different potential future samples is estimated in order to calculate the EVSI using moment matching. Specifically, the mean of these posterior variances are used to rescale the PSA simulations for the known-distribution net benefit. An extension to this method allows us

to calculate the EVSI across different sample sizes using Bayesian non-linear regression to estimate the mean posterior variance for each sample size.

The advantage of this novel methodology is that it places minor restrictions on the structure of the underlying health economic model and data collection process. Specifically, it can be applied provided a Bayesian model can be specified for the model parameters and the sampling distribution of the potential datasets. Chapter 4 presents software that calculates the EVSI based solely on this Bayesian model. The drawbacks of the moment matching method is that it requires additional simulation from the health economic model and can therefore be slower computationally than other EVSI calculation methods. Although throughout this thesis, the moment matching methods are highly competitive in terms of computational time.

Finally, this thesis develops a suite of graphics designed to aid the interpretation of the EVSI, included within the `R` package `EVSI`. These graphics either present the EVSI directly or the ENBS, a function of the EVSI, which calculates the economic value of a proposed trial. Therefore, the ENBS can be used to determine whether a trial should be funded and the optimal design of that trial amongst a set of alternative options. The ENBS has frequently been touted as a potential method for trial design but computational and theoretical barriers have hindered the application of the ENBS in these settings. The graphics within the `EVSI` package should aid researchers to present the ENBS to stakeholders and decision makers to give a tangible understanding of VoI measures by comparing the value and cost of a proposed study.

Future work in this area, therefore, must focus on applying these methods to practical health economic decision models. These applications need to inform trial design and funding applications so practitioners in these areas understand the power of VoI analyses. Now that the computational barriers have largely been overcome, it is important to engage in a dialogue with practitioners and ensure that both the methods and tools will be used in practical health-economic evaluations.

BIBLIOGRAPHY

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

[2] L. Andronis, P. Barton, and S. Bryan. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Prepress Projects Limited*, 2009.

[3] P. Arendt, D. Apley, and W. Chen. A preposterior analysis to predict identifiability in the experimental calibration of computer models. *IIE Transactions*, 48(1):75–88, 2016.

[4] G. Baio. *Bayesian methods in Health Economics*. CRC Press, Boca Raton, FL, 2012.

[5] G. Baio, A. Berardi, and A. Heath. Bayesian cost effectiveness analysis with the R package BCEA, 2016.

[6] G. Baio, A. Berardi, and A. Heath. *BCEA: Bayesian Cost Effectiveness Analysis*, 2016. R package version 2.2-3.

[7] G. Baio and P. Dawid. Probabilistic sensitivity analysis in health economics. *Statistical methods in medical research*, 24(6):615–634, 2011.

[8] M. Ben-Zvi, B. Berkowitz, and S. Kesler. Pre-posterior analysis as a tool for data evaluation: application to aquifer contamination. *Water Resources Management*, 2(1):11–20, 1988.

[9] A. Brennan and S. Kharroubi. Efficient Computation of Partial Expected Value of Sample Information Using Bayesian Approximation. Research Report 560/05, Department of Probability and Statistics, University of Sheffield, UK, 2005.

[10] 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.

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

[12] 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.

[13] A. Briggs, M. Sculpher, and K. Claxton. *Decision modelling for health economic evaluation*. Oxford University Press, Oxford, UK, 2006.

[14] M. Cameletti, R. Ignaccolo, and S. Bande. Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics*, 22:985–996, 2011.

[15] M. Cameletti, F. Lindgren, D. Simpson, and H. Rue. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97(2):109–131, 2013.

[16] J. Campbell, R. McQueen, A. Libby, D. Spackman, J. Carlson, and A. Briggs. Cost-effectiveness Uncertainty Analysis Methods A Comparison of One-way Sensitivity, Analysis of Covariance, and Expected Value of Partial Perfect Information. *Medical Decision Making*, 35(5):596–607, 2015.

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

[18] E. Çetinkaya and A. Thiele. A moment matching approach to log-normal portfolio optimization. *Computational Management Science*, 13(4):501–220, 2016.

[19] K. Chaloner and K. Larntz. Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21(2):191–208, 1989.

[20] W. Chang, J. Cheng, J. Allaire, Y. Xie, and J McPherson. *shiny: Web Application Framework for R*, 2017. R package version 1.0.5.

[21] M. Chen and A. Willan. Value of information methods for assessing a new diagnostic test. *Statistics in medicine*, 33(11):1801–1815, 2014.

[22] P. Ciarlet. *The finite element method for elliptic problems*. Elsevier, New York, NY, 1978.

[23] K. Claxton. Bayesian approaches to the value of information: implications for the regulation of new pharmaceutical. *Health Economics*, 8:269–274, 1999.

[24] K. Claxton. The irrelevance of inference: a decision-making approach to stochastic evaluation of health care technologies. *Journal of Health Economics*, 18:342–364, 1999.

[25] K. Claxton, L. Lacey, and S. Walker. Selecting treatments: a decision theoretic approach. *Journal of the Royal Statistical Society* A, 163:211–225, 2000.

[26] K. Claxton, S. Martin, M. Soares, N. Rice, D. Spackman, S. Hinde, N. Devlin, P. Smith, M. Sculpher, et al. *Methods for the estimation of the NICE cost effectiveness threshold*. 2013.

[27] K. Claxton, P. Neumann, S. Araki, and M. Weinstein. Bayesian Value-Of-Information Analysis. *International Journal of Technology Assessment in Health Care*, 17:38–55, 2001.

[28] K. Claxton, M. Sculpher, C. McCabe, A. Briggs, R. Akehurst, M. Buxton, J. Brazier, and A. O'Hagan. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics*, 14:339–347, 2005.

[29] D Cook and L. Ni. Sufficient dimension reduction via inverse regression. *Journal of the American Statistical Association*, 2005.

[30] R. D. Cook. Fisher lecture: Dimension reduction in regression. *Statistical Science*, pages 1–26, 2007.

[31] R. D. Cook and L. Forzani. Principal fitted components for dimension reduction in regression. *Statistical Science*, 23(4):485–501, 2008.

[32] R. D. Cook and S. Weisberg. Discussion on "Sliced inverse regression for dimension reduction" by K. C. Li. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

[33] N. Cooper, A. Sutton, K. Abrams, D. Turner, and A. Wailoo. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Economics*, 13(3):203–226, 2004.

[34] 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.

[35] Douglas Coyle, Martin J Buxton, and Bernie J O'Brien. Measures of importance for economic analysis based on decision modeling. *Journal of clinical epidemiology*, 56(10):989–997, 2003.

[36] N. Cressie. *Statistics for Spatial Data*. Wiley, 1993.

[37] R. Davis. Gaussian Processes, Encyclopedia of Environmetrics, Section on Stochastic Modeling and Environmental Change. *Wiley, New York*, 2001.

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

[39] P. Diaconis and D. Ylvisaker. Conjugate Priors for Exponential Families. *Ann. Statist.*, 7(2):269–281, 1979.

[40] H. Dong, D. Coyle, and M. Buxton. Value of information analysis for a new technology: Computer-assisted total knee replacement. *International Journal of Technology Assessment in Health Care*, 23, 2007.

[41] S. Durrleman and R. Simon. Flexible regression models with cubic splines. *Statistics in medicine*, 8(5):551–561, 1989.

[42] M. Ebden. Gaussian Processes for regression: A quick introduction. *The Website of Robotics Research Group in Department on Engineering Science, University of Oxford*, 2008.

[43] S. Eckermann, J. Karnon, and A. Willan. The value of value of information. *Pharmacoeconomics*, 28(9):699–709, 2010.

[44] A. Erkanli and R. Soyer. Simulation-based designs for accelerated life tests. *Journal of Statistical Planning and Inference*, 90(2):335–348, 2000.

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

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

[47] J. Felli and G. Hazen. A Bayesian approach to sensitivity analysis. *Health Economics*, 8:263–268, 1999.

[48] L. Feng, J. Korvink, and P. Benner. A fully adaptive scheme for model order reduction based on moment matching. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 5(12):1872–1884, 2015.

[49] E. Fenwick, B. O'Brien, and A. Briggs. Cost-effectiveness acceptability curves–facts, fallacies and frequently asked questions. *Health economics*, 13(5):405–415, 2004.

[50] E. Fenwick, S. Palmer, K. Claxton, M. Sculpher, K. Abrams, and A. Sutton. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. *Medical Decision Making*, 26:480–496, 2006.

[51] 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.

[52] A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1(3):515–534, 2006.

[53] Mark Gibbs and David J.C. MacKay. Efficient implementation of gaussian processes. Technical report, 1997.

[54] T. Hastie and R. Tibshirani. *Generalized additive models*, volume 43. CRC Press, Boca Raton, FL, 1990.

[55] T. Hastie and R. Tibshirani. *Generalized additive models*, volume 43. CRC Press, 1990.

[56] A. Heath and G. Baio. EVSI: A suite of functions for the calculation and presentation of the EVSI. *GitHub repository*, 2017.

[57] A. Heath and G. Baio. An Efficient Calculation Method for the Expected Value of Sample Information: Can we do it? Yes, we can. *Submitted to Value in Health*, -(-):–, 2018.

[58] A. Heath, G. Baio, and R. Hunter. Development of a New Software Tool to Compute the Expected Value of Sample Information: An application to the HomeHealth intervention. *HESG Winter Meeting*, Working Paper(-):–, 2018.

[59] A. Heath, I. Manolopoulou, and G. Baio. Estimating the expected value of partial perfect information in health economic evaluations using integrated nested laplace approximation. *Statistics in medicine*, 35(23):4264–4280, 2016.

[60] 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.

[61] 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. -, -(-):–, 2018.

[62] 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.

[63] N. Higham. Computing the nearest correlation matrix — a problem from finance. *IMA journal of Numerical Analysis*, 22(3):329–343, 2002.

[64] R. Howard. Information Value Theory. In *IEEE Transactions on System Science and Cybernetics*, volume 2:1, pages 22–26. SCC-2, 1966.

[65] Y. Huang and H. Wu. Bayesian experimental design for long-term longitudinal HIV dynamic studies. *Journal of statistical planning and inference*, 138(1):105–113, 2008.

[66] 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.

[67] H. Jalal and F. Alarid-Escudero. A Gaussian Approximation Approach for Value of Information Analysis. *Medical Decision Making*, 38(3):174–188, 2017.

[68] 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.

[69] Z. Jiang, D. Apley, and W. Chen. Surrogate Preposterior Analyses for Predicting and Enhancing Identifiability in Model Calibration. *International Journal for Uncertainty Quantification*, 5(4), 2015.

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

[71] S. Lamb, M. Williams, E. Williamson, S. Gates, E. Withers, S. Mt-Isa, D. Ashby, E. Castelnuovo, M. Underwood, M. Cooke, et al. Managing Injuries of the Neck Trial

(MINT): a randomised controlled trial of treatments for whiplash injuries. *Health Technology Assessment*, 2012.

[72] B. Li and S.i Wang. On directional regression for dimension reduction. *Journal of the American Statistical Association*, 102(479):997–1008, 2007.

[73] K. Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

[74] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

[75] G Loomes and L McKenzie. The use of QALYs in health care decision making. *Social Science and Medicine*, 28:299–308, 1989.

[76] D. MacKay. Introduction to Gaussian Processes. *NATO ASI Series F Computer and Systems Sciences*, 168:133–166, 1998.

[77] J. Madan, A. Ades, M. Price, K. Maitland, J. Jemutai, P. Revill, and N. Welton. Strategies for Efficient Computation of the Expected Value of Partial Perfect Information. *Medical Decision Making*, page 0272989X13514774, 2014.

[78] S. Martino and H. Rue. Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the INLA program. *Department of Mathematical Sciences, NTNU, Norway*, 2009.

[79] S. Martino and H. Rue. Case studies in Bayesian computation using INLA. In *Complex data modeling and computationally intensive statistical methods*, pages 99–114. Springer, 2010.

[80] T. Martins, D. Simpson, F. Lindgren, and H. Rue. Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, 67:68–83, 2013.

[81] C. McCabe, K. Claxton, and A. Culyer. The NICE cost-effectiveness threshold. *Pharmacoeconomics*, 26(9):733–744, 2008.

[82] 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.

[83] D. Meltzer. Addressing uncertainty in medical cost-effectiveness analysis: Implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *Journal of Health Economics*, 20(1):109–129, 2001.

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

[85] P. Müller and G. Parmigiani. Optimal design via curve fitting of Monte Carlo experiments. *Journal of the American Statistical Association*, 90(432):1322–1330, 1995.

[86] J. Oakley. Value of information for complex cost-effectiveness models. *Research Report*, 2002.

[87] National Institute of Health and Care Excellence. Guide to the methods of technology appraisal 2013, 2013.

[88] A. O'Hagan. Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(1), 1978.

[89] A. O'Hagan. Bayes–hermite quadrature. *Journal of statistical planning and inference*, 29(3):245–260, 1991.

[90] A. O'Hagan and J. Stevens. A framework for cost-effectiveness analysis from clinical trial data. *Health Economics*, 10:303–315, 2001.

[91] A. O'Hagan, J. Stevens, and J. Montmartin. Bayesian cost effectiveness analysis from clinical trial data. *Statistics in Medicine*, 20:733–753, 2001.

[92] M. Plummer. *rjags: Bayesian Graphical Models using MCMC*, 2015. R package version 4-4.

[93] M. Plummer et al. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna, Austria, 2003.

[94] J. Pratt, H. Raiffa, and R. Schlaifer. *Introduction to statistical decision theory*. MIT press, 1995.

[95] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

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

[97] C. Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.

[98] C. Rasmussen and C. Williams. Gaussian Processes for Machine Learning., 2006.

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

[100] H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. CRC Press, Boca Raton, FL, 2005.

[101] H. Rue and S. Martino. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of statistical planning and inference*, 137(10):3177–3192, 2007.

[102] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (statistical methodology)*, 71(2):319–392, 2009.

[103] H. Rue, S. Martino, F. Lindgren, D. Simpson, A. Riebler, and E. Krainski. *INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximaxion*, 2014. R package version 0.0-1404466487.

[104] M. Sadatsafavi, N. Bansback, Z. Zafari, M. Najafzadeh, and C. Marra. Need for speed: an efficient algorithm for calculation of single-parameter expected value of partial perfect information. *Value in Health*, 16(2):438–448, 2013.

[105] D. Simpson, F. Lindgren, and H. Rue. Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics*, 1:16–29, 2012.

[106] D. Spiegelhalter, K. Abrams, and J. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley and Sons, Chichester, UK, 2004.

[107] 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.

[108] M. Strong. Partial EVPPI Functions. `http://www.shef.ac.uk/polopoly_fs/1.305039!/file/R_functions.txt`, 2012.

[109] M. Strong, P. Breeze, C. Thomas, and A. Brennan. *SAVI - Sheffield Accelerated Value of Information, Release version 1.013 (2014-12-11)*, 2014.

[110] M. Strong and J. Oakley. An efficient method for computing single-parameter partial expected value of perfect information. *Medical Decision Making*, 33(6):755–766, 2013.

[111] 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.

[112] 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.

[113] 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.

[114] T. Sweeting and S. Kharroubi. Some new formulae for posterior expectations and Bartlett corrections. *Test*, 12(2):497–521, 2003.

[115] P. Thokala, S. Goodacre, M. Ward, J. Penn-Ashman, and G. Perkins. Cost-effectiveness of out-of-hospital continuous positive airway pressure for acute respiratory failure. *Annals of emergency medicine*, 65(5):556–563, 2015.

[116] W. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(1):234–240, 1970.

[117] H. Tuffaha, L. Gordon, and P. Scuffham. Value of Information Analysis Informing Adoption and Research Decisions in a Portfolio of Health Care Interventions. *MDM Policy & Practice*, 1(1), 2016.

[118] B. Weaver, B. Williams, C. Anderson-Cook, D. Higdon, et al. Computational enhancements to Bayesian design of experiments using Gaussian processes. *Bayesian Analysis*, 11(1):191–213, 2016.

[119] M Weinstein. Recent developments in decision-analytic modelling for economic evaluation. *Pharmacoeconomics*, 24(11):1043–1053, 2006.

[120] 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.

[121] A. Willan and A. Briggs. *The statistical analysis of cost-effectiveness data*. John Wiley and Sons, Chichester, UK, 2006.

[122] A. Willan, R. Goeree, and K. Boutis. Value of information methods for planning and analyzing clinical studies optimize decision making and research planning. *Journal of clinical epidemiology*, 65(8):870–876, 2012.

[123] E. Wilson. A Practical Guide to Value of Information Analysis. *PharmacoEconomics*, 33(2):105–121, 2015.

[124] S. Wood. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):95–114, 2003.

[125] S. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36, 2011.

[126] S. Wood, N. Pya, T. Kneib, K. Hornik, M. Lonergan, H. Nilsson, F. Scheipl, and B. Ripley. *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL*, 2016.