A Practical Guide to Value of Information Analysis

Edward C. F. Wilson

Abstract Value of information analysis is a quantitative method to estimate the return on investment in proposed research projects. It can be used in a number of ways. Funders of research may find it useful to rank projects in terms of the expected return on investment from a variety of competing projects. Alternatively, trialists can use the principles to identify the efficient sample size of a proposed study as an alternative to traditional power calculations, and finally, a value of information analysis can be conducted alongside an economic evaluation as a quantitative adjunct to the 'future research' or 'next steps' section of a study write up. The purpose of this paper is to present a brief introduction to the methods, a step-by-step guide to calculation and a discussion of issues that arise in their application to healthcare decision making. Worked examples are provided in the accompanying online appendices as Microsoft Excel spreadsheets.

1 Introduction

Value of information analysis (VoI) is a means of valuing the expected gain from reducing uncertainty through some form of data collection exercise (e.g., a trial or epidemiological study). As such, it is a tool which can be used to assess the cost effectiveness of alternative research projects.

E. C. F. Wilson (✉)
University of Cambridge, Cambridge, UK
e-mail: ed.wilson@medschl.cam.ac.uk

The expected value of a research project is the expected reduction in the probability of making the 'wrong' decision multiplied by the average consequence of being 'wrong' (the 'opportunity loss' of the decision, defined in Sect. 2.1 below). This is compared with the expected cost of the research project. If the expected value exceeds the (expected) cost then the project should be undertaken. If not, then the project should not be undertaken: the (expected) value of the resources consumed by the project exceeds the (expected) value of the information yielded.

VoI is based firmly within a Bayesian statistical framework where probability represents degrees of belief about plausible values for a parameter rather than the long run relative frequency with which an event occurs (as is the case in the frequentist approach). The key concept in Bayesian analysis is the updating of a prior belief about plausible values for a parameter with the support for likely values of that parameter drawn from sampled data (the distribution of which is known as the likelihood function) to form a posterior belief using Bayes theorem [1]. For this reason, Bayesian analysis is sometimes referred to as posterior analysis [2]. VoI requires prediction of the likelihood function conditional on the prior to generate an expected posterior distribution. In lay terms, the results of a data collection exercise (e.g., clinical trial) are predicted based on current knowledge. These are combined with the current knowledge to predict the state of knowledge after the data are collected. It is thus sometimes referred to as preposterior analysis.

The inclusion of VoI as a part of health economic evaluations is increasing [3–12]. This is useful to direct future research effort to where it can achieve the greatest expected return for finite funding. Its primary use is to determine the optimal sample size for a study based on the

marginal gain from an additional trial enrolee compared with the marginal cost. The optimal point is where the marginal cost is equal to the (value of the) marginal gain, a concept directly analogous to the profit maximising condition in the theory of the firm.

The purpose of this paper is to describe briefly the origins of VoI methods and to provide a step-by-step guide to calculation. This manuscript focuses on an analytic approach. However, a numeric (simulation) approach is described in Appendix 1. Spreadsheets with worked examples are also provided as online appendices (see Electronic Supplementary Material [ESM]).

2 Concepts/Descriptive Approach

2.1 The Core Theory

The origins of VoI lie in the work of Raiffa and Schlaifer on statistical decision theory at Harvard [2, 13, 14]. The starting point is that there is some objective function to be maximised, and a choice between courses of action leading to uncertain payoffs with respect to the objective function. It is possible to invest in research to reduce uncertainty in the payoffs, but such information is costly and will thus have a negative impact on the payoff. The question then is whether the decision should be made on current information or whether it is worth investing in additional information to reduce uncertainty before then revisiting the decision.

The payoff can be any outcome such as profit, output or revenue, or broader, less tangible concepts such as happiness, welfare or utility. Likewise, the research can be anything that reduces uncertainty in the payoffs. For example, suppose a medical supplies firm wishes to maximise its profits. It wishes to invest in new manufacturing facilities leading to a much higher level of output allowing it to expand into new markets. However, this will only be profitable if demand is sufficiently high for its product. If demand is lower than expected, sales will be insufficient to make the investment profitable. In this case the objective function is profit, which is uncertain due to uncertainty in demand. The firm can make its decision to invest or not in the new facility now, or it can delay the decision (i.e., maintain the current level of output) and conduct market research to reduce uncertainty in demand and then make its investment decision. The expected cost of the 'delay and research' strategy is the cost of the research itself plus any expected foregone increase in profits had the investment decision been made immediately. The expected value of the strategy is the reduction in expected loss through a reduced probability of making the 'wrong' investment decision.

The same logic also applies to individual decision making. Suppose a utility-maximising consumer is faced with a choice of beers at a bar. The consumer could make the decision as to which to purchase at random. Alternatively he or she could invest in research (request a sample of each) and make the decision based on that new information. The cost of such research is the delayed enjoyment of a beer (assuming zero cost and utility from the sampling process itself), but the benefit is reduced uncertainty as to which is preferred, and hence a higher probability of identifying a preferred beer and thus gaining the most benefit (maximising utility).

In both examples, the principles and questions are the same: does the value of the additional information outweigh its cost? In the former, does the expected profit from a strategy of research followed by investment decision exceed the expected profit from the investment decision now; in the latter, does the expected utility from sampling the range followed by making a decision exceed the expected utility from choosing one at random without sampling.

The key measurements in VoI are the expected value of perfect information (EVPI), expected value of sample information (EVSI) and the expected net gain of sampling (ENGS, sometimes termed the expected net benefit of sampling, ENBS). The expected value of perfect parameter information (EVPPI) is also sometimes defined. This is the value of eliminating uncertainty in one or more input parameter(s) of the objective function. (Note the EVPPI is also sometimes termed the expected value of partial perfect information).

Where there are only two courses of action, A or B, the decision is most easily represented by calculating the incremental expected payoff of one option compared with the other; that is, the expected payoff with option B less the expected payoff with A. The expected incremental net payoff (or incremental net benefit) and its associated uncertainty can be plotted as per Fig. 1. A cash value (e.g., profit) is used for the payoff in this example, but the principles are the same whether the payoff is cash, utility or some other metric. The incremental payoff is referred to from hereon as the incremental net benefit (INB), and denoted 'ΔB' in subsequent equations.

Based on current information, the expected INB is positive (+£300 in Fig. 1). The decision should therefore be in favour of option B. However, due to uncertainty there is a probability that the decision is wrong, represented by the shaded area in Fig. 1. If it turns out that the INB is actually say, −£250, the wrong decision will have been made: the payoff would have been £250 higher had the decision been to go with option A; the loss (termed the opportunity loss) is therefore £250. Likewise, if the INB was actually −£500, the opportunity loss is £500.
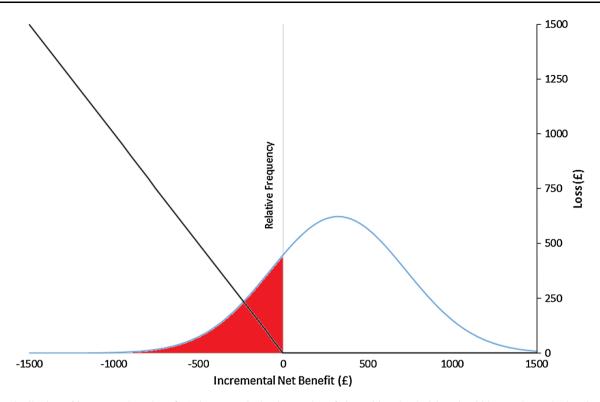
Fig. 1 Distribution of incremental net benefit (primary vertical axis) and loss function (secondary vertical axis). The distribution of incremental net benefit (ΔB) is indicated by the *blue curve*. The loss function is indicated by the *black line*. As mean incremental net benefit is positive the decision should be to adopt. The loss is zero to the right of the origin, and equal to $-\Delta B$ to the left. The proportion of the area under the *curve shaded* is the probability of a loss given a decision to adopt

The opportunity loss can therefore be plotted in relation to a secondary *y*-axis as a $-45°$ line from $-\infty$ to zero (Fig. 1). If it turns out that INB is, say, $+£100$, or indeed any positive value, there is no opportunity loss as the decision to go with option B was the correct decision. The loss function therefore kinks at the origin and coincides with the *x*-axis at values greater than zero.

In simple terms, the probability of being 'wrong' multiplied by the average consequence of being wrong (the opportunity loss) is the expected loss associated with uncertainty, or equivalently the expected gain from eliminating uncertainty, which is the EVPI.

This logic can be demonstrated most clearly with a discrete approximation. In Fig. 2a, the continuous distribution shown in Fig. 1 is approximated by two possible discrete payoffs: a 23 % probability of incurring a loss of (approximately) £500, and a 77 % probability of a gain of (approximately) £500. The expected payoff (i.e. INB) is therefore $0.23 \times -500 + 0.77 \times 500 = £270$, and the expected loss $0.23 \times 500 = £115$.

In Fig. 2b, the same decision problem is divided into four discrete payoffs of (approximately) $-£750$, $-£250$, $+£250$ and $+£750$, with associated probabilities of 2.3, 20.4, 46.5 and 30.8 %, respectively. The expected INB is therefore $0.023 \times -750 + 0.204 \times -250 +$ $0.465 \times 250 + 0.308 \times 750 = £279$, with an expected loss of $0.023 \times 750 + 0.204 \times 250 = £68$. In Fig. 2c, the problem is further subdivided, yielding an expected INB of £298 and expected loss of £52. Continual subdivision of the problem until each discrete column is an 'infinitesimal strip' equates to the continuous case as illustrated in Fig. 1 (an expected value of £300 and expected loss of £52).

Suppose some research activity can be undertaken which will reduce uncertainty in the INB (i.e., reduce decision uncertainty). The results of this research can be predicted with the likelihood function: the most likely value of the sampled INB is the prior INB. Given knowledge of the standard deviation of INB, the expected reduction in standard error from a study of a given size can be calculated when the prior is combined with the predicted sample results. This will 'tighten' the distribution and thus reduce the probability of making the wrong decision (proportion of the probability mass represented by the shaded area in Fig. 3), hence reducing the expected loss associated with uncertainty. (Note the pre-posterior mean will always equal the prior mean as the most likely value for the sample mean is the prior mean).

The expected reduction in expected loss is the expected gain from that sample information, or the EVSI.
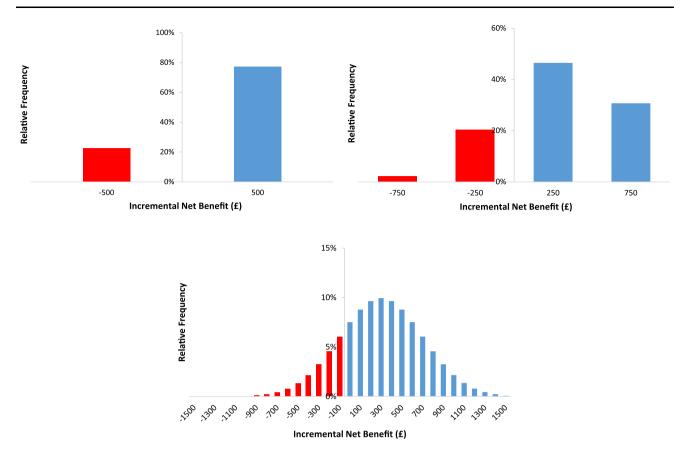
Fig. 2 **a**, **b** and **c** Discrete approximation for calculating the expected loss associated with a decision with uncertain payoffs
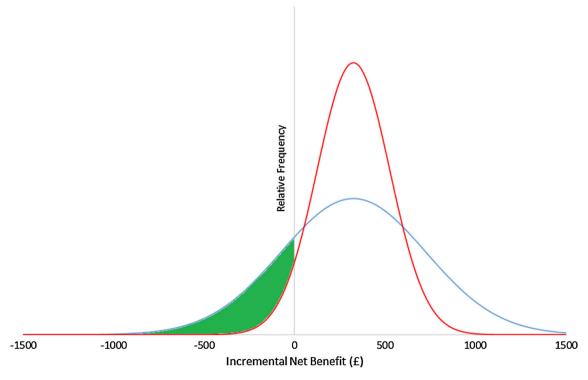

Fig. 3 Prior and predicted posterior distribution of incremental net benefit. The *blue line* indicates the prior distribution of incremental net benefit (ΔB), with the *red line* indicating the predicted posterior. The expected reduction in probability of a loss is equal to the shaded proportion of the area under the prior distribution function

A small research study will yield a small EVSI, whilst a larger study will yield a bigger EVSI. But a larger study will also cost more than a smaller one. The difference between the EVSI and the cost of the study is the ENGS. The sample size that maximises the ENGS by definition maximises the expected return on investment and is the optimal size for a research study.

2.2 Application to Decision Making in the Healthcare Field

The principles were first adapted to the healthcare field by Thompson [15], with substantial development undertaken by, among others, Claxton, Briggs, Willan and Eckermann [16–18]. VoI is probably most usefully considered as a step in the iterative approach to decision making and research [19–23]. This comprises firstly defining the decision problem followed by systematic review of all relevant evidence, which is then combined together in a decision model. Point estimate results of the decision model are used to inform the adoption decision whilst decision uncertainty is used to inform the research decision. If new research is deemed worthwhile, it should be undertaken and the results fed back into the systematic review, at which point the cycle is repeated. Of importance in this approach is the existence of two distinct decisions: the adoption decision and the research decision. As stated above, the adoption decision should be made on expected values alone, whilst uncertainty is used to inform whether it is worth obtaining additional information to reduce that uncertainty.

For example, suppose a new treatment were proposed for a disease to replace existing therapy. The decision problem is whether to adopt the new treatment in place of old. Economic theory would suggest this should be made on the basis of whether it represents a net gain to society, taking into account the opportunity cost of the new treatment (that is, the value of health foregone elsewhere in the system to make way for the new treatment). This is measured by the incremental net monetary benefit of the new treatment, and is simply a rearrangement of the incremental cost-effectiveness ratio decision rule (Eq. 1) [24]. This becomes the objective function to be maximised (Eq. 2). Note that the equation can also be expressed in terms of the incremental net health benefit by dividing both sides of the equation by λ (the value placed on a unit of health gain), but net monetary benefit is more practical to work with (the former leads to divide by zero errors when $\lambda = 0$).

At this point, it is not specified whether the estimate of INB is derived from a single trial or from a decision model based on a synthesis of all relevant evidence. In order to fully reflect current decision uncertainty, the latter is preferable. However, depending on the decision question

and state of current knowledge, a single clinical trial with piggybacked economic evaluation may be an appropriate source of data: Eq. 2 shows INB as a function of incremental cost and outcome alone (as well as the value of a unit of outcome, λ) without specifying how those two parameters are generated.

3 Step-by-Step Calculation

There are two methods by which the VoI statistics can be calculated: analytically, requiring assumptions of normality amongst parameters, and numerically (via simulation), which, whilst relaxing the normality assumptions (allowing alternative parametric forms), can be very burdensome requiring many hours of computer processing time to calculate. The analytic method is most frequently performed on economic evaluations conducted alongside clinical trials, whilst the numeric approach is more often associated with decision models, although in principle either can be applied to either situation. A step-by-step approach to the analytic approach follows, with a description of the simulation approach in Appendix 1. Spreadsheets with the calculations are provided in the ESM, Appendices 2 and 3.

The analytic solution illustrated here assumes mean INB is a simple linear combination of incremental mean cost and outcomes as per Eq. 2. Outcomes are assumed to be measured in quality-adjusted life-years (QALYs) throughout and a threshold of £20,000 per QALY gained is assumed unless otherwise stated. Where sample data provide the source of the priors, calculation of mean and variance of mean INB and its components are as follows:

Individual observations on cost and QALYs are denoted with lower-case letters, and means with upper-case (Eqs. 3, 4), with sample variances and covariance (denoted with lower-case letters) in Eqs. 5–7. The net benefit of patient i in arm j is defined as the value of the QALYs gained by that patient less the cost (Eq. 8). Mean net benefit in arm j can be defined either as the sum of per patient net benefit divided by the number of observations or as the difference between the value of mean QALYs and cost (Eq. 9). Likewise, the sample variance of net benefit in arm j can be defined either from the individual observations on b, or as the sum of the sample variances of QALYs and cost less twice the covariance (Eq. 10).

Variances of means (denoted with capital letters) are equal to the sample variances divided by the sample size (Eqs. 11–14). Note the square root of the sample variance is the standard deviation (a measure of the dispersion of individual observations around the mean) and the square root of the variance of the mean is the standard error (a measure of uncertainty in the estimate of the mean). As per Eq. 10, the variance of mean net benefit can be expressed

either as the sample variance of net benefit divided by the sample size, or the sum of the variances less twice the covariance of mean QALYs and cost (Eq. 14).

Mean incremental cost and QALYs are simply the difference between the cost and QALYs in each arm, respectively (Eqs. 15–16). INB can be expressed likewise (Eq. 17), or as previously defined in Eq. 2. The variances of mean incremental cost and QALYs and the covariance between the mean increments are simply the sum of the respective (co)variances in each arm (Eqs. 18–20). The variance of mean INB can be expressed either as the sum of the variances of mean net benefit, or as the sum of each component (QALYS and cost) less twice the covariance (Eq. 21). Noting that the correlation coefficient between mean incremental cost and QALYs is defined as the covariance of the means divided by the product of the standard errors (Eq. 22), Eq. 21 can be re-written as per Eq. 23. (This is a more useful expression for calculating the EVPPI, see below). The parameters defined in Eqs. 15–23 form the respective priors, denoted with the subscript '$_0$' (Eq. 24).

3.1 Equation Set 1a: Mean Incremental Net Benefit

$$\frac{C_T - C_C}{E_T - E_C} = \frac{\Delta C}{\Delta E} \leq \lambda \tag{1}$$

where: C_j = mean cost per patient of intervention j, E_j = mean outcome (e.g., QALYs gained) per patient from intervention j, $\Delta X = X_2 - X_1$, λ = value placed on/maximum willingness to pay for a unit of outcome.

$$\lambda \Delta E - \Delta C = \Delta B \tag{2}$$

3.2 Equation Set 1b: Derivation of Prior Estimates of Means and Variances of Means from Sample Data

3.2.1 Sample Means and Sample Variances/Covariance by Treatment Arm

$$C_j = \frac{\sum_{i=1}^{n_j} c_{i,j}}{n_j} \tag{3}$$

$$E_j = \frac{\sum_{i=1}^{n_j} e_{i,j}}{n_j} \tag{4}$$

$$v(c_j) = \frac{\sum_{i=1}^{n_j} (c_{i,j} - C_j)^2}{(n_j - 1)} \tag{5}$$

$$v(e_j) = \frac{\sum_{i=1}^{n_j} (e_{i,j} - E_j)^2}{(n_j - 1)} \tag{6}$$

$$\text{Cov}(e_j, c_j) = \frac{\sum_{i=1}^{n_j} (e_{i,j} - E_j)(c_{i,j} - C_j)}{(n_j - 1)} \tag{7}$$

where: $c_{i,j}$ = cost of patient i in arm j ($j = T$, treatment or C, control), $e_{i,j}$ = QALYs gained by patient i in arm j, C_j = mean cost per patient in arm j, E_j = mean QALYs per patient in arm j, n_j = sample size in arm j

$$b_{i,j} = \lambda e_{i,j} - c_{i,j} \tag{8}$$

$$B_j = \frac{\sum_{i=1}^{n_j} b_{i,j}}{n_j} = \lambda E_j - C_j \tag{9}$$

$$v(b_j) = \frac{\sum_{i=1}^{n_j} (b_{i,j} - B_j)^2}{(n_j - 1)}$$
$$= \lambda^2 v(e_j) + v(c_j) - 2\lambda \text{Cov}(e_j, c_j) \tag{10}$$

3.2.2 Variance of Means by Treatment Arm

$$v(C_j) = \frac{v(c_j)}{n_j} \tag{11}$$

$$v(E_j) = \frac{v(e_j)}{n_j} \tag{12}$$

$$\text{Cov}(E_j, C_j) = \frac{\text{Cov}(e_j, c_j)}{n_j} \tag{13}$$

$$v(B_j) = \frac{v(b_j)}{n_j} = \lambda^2 v(E_j) + v(C_j) - 2\lambda \text{Cov}(E_j, C_j) \tag{14}$$

3.2.3 Increments: Means and Variance of Mean

$$\Delta C = C_T - C_C \tag{15}$$

$$\Delta E = E_T - E_C \tag{16}$$

$$\Delta B = B_T - B_C \tag{17}$$

$$v(\Delta C) = v(C_T) + v(C_C) \tag{18}$$

$$v(\Delta E) = v(E_T) + v(E_C) \tag{19}$$

$$\text{Cov}(\Delta E, \Delta C) = \text{Cov}(E_T, C_T) + \text{Cov}(E_C, C_C) \tag{20}$$

$$v(\Delta B) = v(B_T) + v(B_C)$$
$$= \lambda^2 v(\Delta E) + v(\Delta C) - 2\lambda \text{Cov}(\Delta E, \Delta C) \tag{21}$$

$$\rho(\Delta E, \Delta C) = \frac{\text{Cov}(\Delta E, \Delta C)}{\sqrt{v(\Delta E)}\sqrt{v(\Delta C)}} \tag{22}$$

$$v(\Delta B) = \lambda^2 v(\Delta E) + v(\Delta C)$$
$$- 2\lambda \rho(\Delta E, \Delta C)\sqrt{v(\Delta E)}\sqrt{v(\Delta C)} \tag{23}$$

$$X_0 = X \tag{24}$$

where $X = \Delta C$, ΔE, ΔB, $v(\Delta C)$, $v(\Delta E)$, $\text{Cov}(\Delta E, \Delta C)$, $v(\Delta B)$, $\rho(\Delta E, \Delta C)$

3.3 Expected Value of Perfect Information

The EVPI is calculated as per Eq. 25. Note, if mean INB (ΔB) is positive then the indicator function in Eq. 25 reduces the second term in the equation to zero, and the EVPI is $\int_{-\infty}^{0} -\Delta B f_0(\Delta B) d_{\Delta B}$: the integral is from $-\infty$ to zero because if the 'true' value of b is greater than zero, then the correct decision has been made and there is thus no opportunity loss. However, if the 'true' value of b is actually negative, then the wrong decision has been made, and the loss is $-\Delta B$.

The per-patient EVPI is multiplied by N, the total present and (discounted) future population who could benefit from the information. Depending on the disease, this may comprise the current prevalence, plus the incidence over an 'appropriate' time horizon, discounted at an 'appropriate' rate (Eq. 26). If INB is assumed to be normally distributed, the EVPI can be estimated via the unit normal linear loss integral (UNLLI, or standardised loss, denoted L_{N^*}; Eq. 27) [2, 18]. Briefly, the standardised loss evaluated at z is the difference between y and z (where $y > z$) multiplied by the probability of observing that difference in a standard normal variable, summed over all possible values of y from z to ∞ (this is the process illustrated in Fig. 2 but for a standard normal variable). Equation 28 rearranges this into a more readily computable form, where z is the absolute normalised mean INB, $\frac{|\Delta B_0|}{\sqrt{v(\Delta B)_0}}$. The standardised loss is a function of this, the standard normal probability density function, $\phi(z)$ and cumulative distribution function, $\Phi(z)$ (Eqs. 29–30). A good non-technical explanation of loss functions is provided in the Appendix to Cachon and Terwiesch [25].

3.3.1 Equation Set 2: Expected Value of Perfect Information

$$\text{EVPI}_0 = N \left[I\{\Delta B_0 \geq 0\} \int_{-\infty}^{0} -\Delta B f_0(\Delta B) d_{\Delta B} \right.$$
$$\left. + I\{\Delta B_0 < 0\} \int_{0}^{\infty} \Delta B f_0(\Delta B) d_{\Delta B} \right] \quad (25)$$

where N = beneficial population:

$$N = P_0 + \sum_{t=0}^{T} \frac{I_t}{(1+r)^t} \quad (26)$$

P_0 = prevalent population at time $t = 0$,
I_t = incident population at time t,
r = discount rate,
$I\{.\}$ is the indicator function which returns 1 if the condition $\{\}$ is satisfied, otherwise 0,
$f_0(\Delta B)$ = prior density function of ΔB.

$$\text{EVPI}_0 = N \sqrt{v(\Delta B)_0} L_{N^*} \left(\Delta B_0, \sqrt{v(\Delta B)_0} \right) \quad (27)$$

where:

$$L_{N^*}\left(\Delta B_0, \sqrt{v(\Delta B)_0} \right) = \int_{\frac{|\Delta B_0|}{\sqrt{v(\Delta B)_0}}}^{\infty} \left(y - \frac{|\Delta B_0|}{\sqrt{v(\Delta B)_0}} \right) \phi(y) dy$$
$$= \phi\left(\frac{|\Delta B_0|}{\sqrt{v(\Delta B)_0}} \right) - \frac{|\Delta B_0|}{\sqrt{v(\Delta B)_0}} \left[\Phi\left(-\frac{|\Delta B_0|}{\sqrt{v(\Delta B)_0}} \right) - I\{\Delta B_0 < 0\} \right] \quad (28)$$

$\phi(z)$ = standard normal pdf evaluated at z (Eq. 29)
$\Phi(z)$ = standard normal cdf evaluated from $-\infty$ to z (Eq. 30)

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{z^2}{2}\right)} \quad (29)$$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\left(\frac{z^2}{2}\right)} \quad (30)$$

3.3.2 Example

Suppose a trial-based economic evaluation comparing Control with Treatment yielded the following:

Mean INB ΔB_0 = £1,000.

Standard Error of Mean INB $\sqrt{v(\Delta B)_0}$ = £1,500.

Further suppose the present and future beneficial population totals 10,000 patients. As ΔB_0 is greater than zero, the decision would be to adopt Treatment in place of Control. The EVPI would establish whether there could be a case for repeating the trial to reduce decision uncertainty, $v(\Delta B)_0$.

Therefore the EVPI (Eq. 27) is:

$$\text{EVPI}_0 = 10,000 \times 1,500 \left(\phi\left(\frac{|1,000|}{1,500} \right) \right.$$
$$\left. - \frac{|1,000|}{1,500} \left[\Phi\left(-\frac{|1,000|}{1,500} \right) - I\{1,000 < 0\} \right] \right)$$
$$= 10,000 \times 1,500 \times (0.3194 - 0.6667$$
$$\times [0.2525 - 0]) = £2.267 \text{ m}$$

The code to implement this in Microsoft Excel is provided in the ESM, Appendix 2, Sheet 1, Cells B2:D9.

3.4 Expected Value of Perfect Parameter Information

The EVPPI can be estimated by assessing the impact of reducing the standard error of a particular parameter to zero on the reduction in standard error of overall INB. In other words, the EVPPI is the (expected) reduction in expected loss from the reduction in overall decision uncertainty attributable to eliminating uncertainty in a particular parameter.

For example, if ΔC were to be known with certainty, then the posterior variance of ΔC, $v(\Delta C)_1$ would equal 0. Noting that $v(\Delta E)_1 = v(\Delta E)_0$ and $\rho(\Delta E, \Delta C)_1 = \rho(\Delta E, \Delta C)_0$, the posterior variance of ΔB, denoted $v(\Delta B)_1$, is simply the prior estimate of the variance of ΔE (denoted ΔE_0 and converted into monetary units with λ^2, Eqs. 30, 31). The (expected) reduction in variance of ΔB conditional on $v(\Delta C)_1 = 0$, denoted $v(\Delta B)_{s|v(\Delta C)_1=0}$, is therefore the difference between prior and (expected) posterior variance of ΔB (Eq. 32) and the EVPPI calculated as per Eq. 33 (compare this with Eq. 27, where $v(\Delta B)_{s|v(\Delta C)_1=0}$ is substituted in place of $v(\Delta B)_0$). The equivalent is true for the value of eliminating uncertainty in ΔE, where the reduction in uncertainty is as per Eq. 34.

3.4.1 Equation Set 3: Expected Value of Perfect Parameter Information

$$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1$$
$$- \lambda 2\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}\rho(\Delta E, \Delta C)_1 \quad (31)$$

$$\therefore v(\Delta B)_{1|v(\Delta C)_1=0} = \lambda^2 v(\Delta E)_1 + 0 - 0 \quad (32)$$

where: $v(X)_1$ = predicted posterior (i.e. preposterior) variance of mean of X

$$v(\Delta B)_{s|v(\Delta C)_1=0} = v(\Delta B)_0 - v(\Delta B)_{1|v(\Delta C)_1=0}$$
$$= \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - \lambda 2\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0}\rho(\Delta E, \Delta C)_0$$
$$- \lambda^2 v(\Delta E)_0 = v(\Delta C)_0 - \lambda 2\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0}\rho(\Delta E, \Delta C)_0$$
$$(33)$$

$$EVPPI_{\Delta C} = N\sqrt{v(\Delta B)_{s|v(\Delta C)_1=0}}L_{N*}\left(b_0, \sqrt{v(\Delta B)_{s|v(\Delta C)_1=0}}\right)$$
$$(34)$$

where L_{N*} is calculated as per Eq. 28

$$v(\Delta B)_{s|v(\Delta E)_1=0} = v(\Delta B)_0 - v(\Delta B)_{1|v(\Delta E)_1=0}$$
$$= \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - \lambda 2\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0}\rho(\Delta E, \Delta C)_0 - v(\Delta C)_0$$
$$= \lambda^2 v(\Delta E)_0 - \lambda 2\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0}\rho(\Delta E, \Delta C)_0$$
$$(35)$$

3.4.2 Example

Continuing the previous example, suppose the standard error of INB is a function of the standard errors of ΔE and ΔC as per Eq. 23, with a threshold of $\lambda = £20,000$:

Mean INB $\Delta B_0 = £1,000$.

Standard error of mean incremental QALYs $\sqrt{v(\Delta E)_0} = 0.036$.

Standard error of mean incremental cost $\sqrt{v(\Delta C)_0} = £1,000$.

Correlation coefficient between mean incremental QALYs and cost $\rho(\Delta E, \Delta C)_0 = -0.5$.

The standard error of mean INB is now calculated as (Eq. 23):

$$\sqrt{v(\Delta B)_0}$$
$$= \sqrt{20,000^2 \times 0.036^2 + 1,000^2 - 20,000 \times 2 \times 0.036 \times 1,000 \times -0.5}$$
$$= £1,500$$

If uncertainty in ΔC were eliminated, then $v(\Delta C)_1 = 0$ by definition. Therefore as per Eq. 31, $\sqrt{v(\Delta B)_1} = \sqrt{\lambda^2 v(\Delta E)_0} = \sqrt{20,000^2 \times 0.036^2} = £724.75$.

The overall reduction in the standard error of INB from elimination of uncertainty in ΔC is thus (Eq. 32):

$$\sqrt{v(\Delta B)_{s|v(\Delta C)_1=0}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_{1|v(\Delta C)_1=0}}$$
$$= \sqrt{1,500^2 - 724.75^2} = £1,313$$

The EVPPI is then (Eq. 33):

$$EVPPI_{\Delta C} = 10,000 \times 1313\left(\phi\left(\frac{|1,000|}{1,313}\right)\right.$$
$$- \frac{|1,000|}{1,313}\left[\Phi\left(-\frac{|1,000|}{1,313}\right) - I\{1,000 < 0\}\right]\right)$$
$$= 10,000 \times 1,313 \times (0.2985 - 0.761$$
$$\times [0.2232 - 0]) = £1.689\,m$$

Note the calculations presented here are subject to rounding errors: ESM Appendix 2, Sheet 1, Cells G2:I21 provides relevant Excel code and precise figures.

3.5 Expected Value of Sample Information

The predicted posterior EVPI, $EVPI_1$, is uncertain as it is conditional on the trial data, which are unknown. Therefore the expected $EVPI_1$ is the $EVPI_1$ associated with a particular sample result (denoted ΔB_s), multiplied by the probability of observing that result, summed over all possible values of ΔB_s (Eq. 35). The predicted distribution of ΔB_s, denoted $\hat{f}(\Delta B_s)$, is the likelihood function for different values of ΔB_s. The EVSI is thus the difference between prior EVPI and expected posterior EVPI, which is then multiplied by the patient population, N, less those enrolled in the study, $2n_s$ as (depending on the nature of the disease) they cannot benefit from the information (Eq. 36).

Willan and Pinto [26] provide a comprehensive approach to calculating the EVSI. A simpler notation can be derived from Eq. 27 replacing $\sqrt{v(\Delta B)_0}$ with the reduction in standard error of INB from a trial of sample size n_s per arm, $\sqrt{v(\Delta B)_{s,n}}$ and the potentially beneficial

population is the total population less those enrolled in the study (Eq. 37) [2]. Thus, $v(\Delta B)_{s,n}$ is the difference between prior and (expected) posterior variance of mean INB and is calculated as per Eq. 38. n_0 is the prior sample size which may be known where there are actual prior data or inferred by rearranging Eq. 14 (i.e., the ratio of the sample variance and variance of the mean).

Where $v(b_T)$ and $v(b_C)$ and hence $v(\Delta b)$ (Eq. 39) are unknown, appropriate estimates may be obtained from the literature in related disease areas or from expert opinion, as is common practice when undertaking conventional power calculations.

3.5.1 Equation Set 4: Expected Value of Sample Information

$$E(\text{EVPI}_1) = \int_{-\infty}^{\infty} \text{EVPI}_1 \hat{f}(\Delta B_s) d_{\Delta B_s} \qquad (36)$$

$$E_{\Delta B_s}(\text{EVSI}(n_s, \Delta B_s)) = (N - 2n_s)[\text{EVPI}_0 - E(\text{EVPI}_1)] \qquad (37)$$

where: n_s = number of observations per arm.

$$\text{EVSI}_n = (N - 2n_s)\sqrt{v(\Delta B)_{s,n}} L_{N*}\left(\Delta B_0, \sqrt{v(\Delta Bb)_{s,n}}\right) \qquad (38)$$

where L_{N*} is calculated as per Eq. 28, substituting $v(\Delta B)_{s,n}$ in place of $v(\Delta B)_0$.

$$
\begin{aligned}
v(\Delta B)_{s,n} &= v(\Delta B)_0 - v(\Delta B)_1 \\
&= v(\Delta B)_0 - \frac{v(\Delta b)}{n_0 + n_s} \\
&= v(\Delta B)_0 - \left(\frac{v(\Delta b)}{\frac{v(\Delta b)}{v(\Delta B)_0} + n_s}\right) \qquad (39) \\
&= v(\Delta B)_0 - \left(\frac{1}{v(\Delta B)_0} + \frac{n_s}{v(\Delta b)}\right)^{-1}
\end{aligned}
$$

where: n_0 is the sample size associated with the prior $v(\Delta b)$ is the sum of the sample variances of b in each arm:

$$v(\Delta b) = v(b_T) + v(b_C) \qquad (40)$$

where $v(b_T)$ and $v(b_C)$ are calculated as per Eq. 10.

3.5.2 Example

Continuing the example above, suppose $v(b_T) = v(b_C) = $ £50,000,000, thus $v(\Delta b) = $ £100,000,000 (obtained either from previous studies as per Eq. 39 or via elicitation as described above). Let $\lambda = $ £20,000 and suppose a study of sample size $n = 100$ per arm is proposed. First calculate the (expected) reduction in variance of mean INB (Eq. 38):

$$
\begin{aligned}
\sqrt{v(\Delta B)_{s,100}} &= 1,500^2 - \left(\frac{1}{1,500^2} + \frac{100}{100,000,000}\right)^{-1} \\
&= \text{£}1,248
\end{aligned}
$$

The EVSI is then the unit normal loss multiplied by the reduction in standard error and by the beneficial population as shown previously (Eq. 37):

$$
\begin{aligned}
\text{EVSI} &= (10,000 - 2 \times 100) \times 1,248 \\
&\quad \times L_{N*}(1,000, 1,248) \\
&= (10,000 - 2 \times 100) \times 1,248 \left(\phi\left(\frac{|1,000|}{1,248}\right)\right. \\
&\quad \left. - \frac{|1,000|}{1,248}\left[\Phi\left(-\frac{|1,000|}{1,248}\right) - 0\right]\right) \\
&= 9,800 \times 1,248 \times 0.1202 = \text{£}1.467\,\text{m}
\end{aligned}
$$

As with the previous examples, the numbers presented here are subject to rounding errors. Full working and Excel code is in ESM, Appendix 2, Sheet 1, Cells B12:D20.

3.6 Expected Net Gain of Sampling

The expected net gain of sampling is the expected gain from the trial (i.e., EVSI) less the cost of sampling (total cost [TC], Eqs. 40, 41). Note that both the EVSI and TC (and thus ENGS) are functions of n. The calculations should be repeated for a wide range of values of n_s, and the optimal n_s (denoted $n*$) is that which maximises the ENGS.

3.6.1 Equation Set 5: Expected Net Gain of Sampling

$$\text{TC}_n = C_f + 2n_s C_v + n_s|\Delta B_0| \qquad (41)$$

where C_f is the fixed cost of sampling and C_v is variable (per patient) cost of sampling

$$\text{ENGS}_n = \text{EVSI}_n - \text{TC}_n \qquad (42)$$

3.6.2 Example

Suppose the fixed costs of a trial totalled £50,000 and a variable cost of £250 per patient enrolled. A trial of size $n = 100$ per arm would therefore cost (Eq. 40):

$$
\begin{aligned}
\text{TC}_{n=100} &= 50,000 + 2 \times 100 \times 250 + 100 \times 1,000 \\
&= \text{£}200,000
\end{aligned}
$$

The ENGS of a trial of 100 patients in each arm is thus £1.467 m −£0.2 m = £1.267 m. As this is greater than zero, this trial would be worthwhile; however, the calculations should be repeated for a range of values of n_s to identify the ENGS-maximising n_s (denoted $n*$). Figure 4 shows the ENGS for a range of sample sizes,
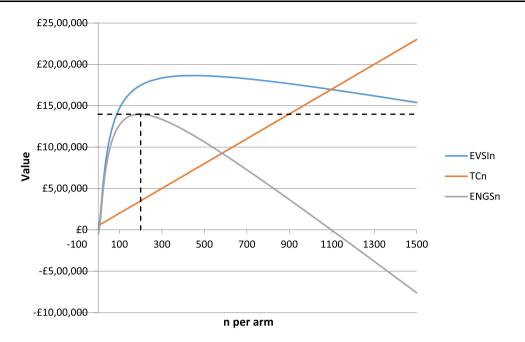
Fig. 4 Expected value of sample information, total cost and expected net gain of sampling by sample size. *EVSI_n* Expected value of sample information of a study of sample size *n* per arm, TC_n = total cost of a study of sample size *n* per arm, $ENGS_n$ = expected net gain of sampling from a study of sample size *n* per arm

identifying the optimum at approximately 200 patients per arm (see ESM, Appendix 2 for calculations).

4 Discussion and Conclusion

This paper aims to provide a 'hands on' guide to using VoI, providing a working template to assist readers in conducting their own analyses. The worked examples show the analytic approach, whilst the numeric approach is detailed in Appendix 1. Both have their respective advantages and disadvantages. The major advantage of the analytic approach is that it is fast to calculate, and is not subject to random 'noise' (Monte Carlo error) intrinsic in simulation methods. The major disadvantage is the assumption of normally distributed parameters. Conversely, the advantage of the numeric approach is its flexibility with regards to the distributional form of both input and output parameters; however, it can be time consuming to run sufficient simulations in order to minimise Monte Carlo errors. Comparisons of the results of the analytic and numeric approaches to the same decision problem would be a useful addition to the literature.

Steuten et al. [27] recently conducted a systematic review of the literature covering both development of methods and application of VoI. The review identified a roughly 50/50 split between methodological and applied examples. Amongst the applications, most succeeded in calculating the EVPI and/or EVPPI, but very few went on

to calculate the EVSI. A possible reason for this could be the computational burden, with some analyses requiring weeks of computer processing time. Steuten and colleagues [27] acknowledge a number of studies concerned with efficient computation of EVSI, and conclude with a recommendation that future research should focus on making VoI applicable to the needs of decision makers.

There are a number of methodological challenges that have arisen in adapting VoI to the healthcare sector, the most important of which is defining the scope of the benefits from the proposed trial. In the case of a firm conducting market research, the expected net benefit of the research is simply the net impact on expected profit. However, healthcare applications usually seek to inform policy decisions for the benefit of a population. Most economic evaluations express the INB on a per-patient scale. Thus, the EVPI and EVSI are also expressed per patient. To estimate the gain to the health economy, the EVPI and EVSI must be multiplied by the patient population. However defining this is far from straightforward. Those who could potentially benefit from the information include the prevalent cohort with the disease in question and/or the future incident population. Whilst it may be possible to estimate the future incidence and prevalence of the disease with a reasonable degree of accuracy, the time horizon over which the incidence should be calculated is unclear. Most studies use 10–20 years as a de facto standard (and discount the benefit to future populations at the prevailing rate), but without any clear justification [28].

This is of concern as the VoI statistics can be highly sensitive to the time horizon.

After determining the relevant prevalence and incidence, it is argued that patients who participate in the study will not benefit from the information yielded (although this depends on whether the condition is acute or chronic [29]). Therefore, the beneficial population is usually reduced by the numbers of patients enrolled in a study [18, 26]. Likewise, patients enrolled in the 'inferior' arm of a study incur an opportunity cost equal to the foregone INB per patient (which is usually added to the total cost of conducting the study). The impact of these issues on the overall value of information depends on the size of the patient population relative to those enrolled in the trial. For a common disease such as asthma or diabetes, trial enrolees will comprise a very small proportion of the total population. However, for rarer diseases, accounting for the opportunity cost of trial enrolees may affect the optimal sample size calculations substantially.

A number of other issues in adapting VoI to the healthcare setting relate to the independence (or lack thereof) of the adoption and research decisions. Whilst conceptually separate, they are not independent of one another as (i) if the adoption decision is delayed whilst new research is underway, there will be an opportunity cost to those who could have benefited if the technology does indeed have a positive INB [30] (and vice versa: if the technology actually has a negative INB and it is adopted with a review of the decision following further research then patients would have been better off with the old treatment); and (ii) if there are considerable costs associated with reversing a decision [17]; for example, retraining of staff or costly conversion of facilities to other uses ([31] cited in [17]).

The former issue has the potential to dramatically reduce the expected value of information: if the time horizon for the analysis is 10 years, but it takes 5 years for a proposed study to be conducted and disseminated, the value of sample information could be (more than) halved. The latter issue can be addressed by adopting an option pricing approach borrowed from financial economics, where the expected value of a strategy to reject with the option to accept pending further evidence compared with a strategy of immediate adoption or rejection is calculated [30]. This requires adding in the (expected, present value) cost of future reversal to the cost of a strategy of immediate adoption [17], and comparing the net benefit of this with one of delay followed by investment.

The final issue relates to the nature of information as a public good: once in the public domain it is non-rival and non-excludable meaning consumption by one individual or group neither diminishes consumption by another, nor can that individual or group prevent another from consuming it. Ignoring other potential benefits to an economy from research (e.g., employment maintenance and prestige), this would lead to free riding as there is no reason for one jurisdiction (e.g., a state research funder) to pay for research when another can do so. Therefore, whilst the EVSI may suggest a particular study should be carried out, it may be strategically optimal to wait for another jurisdiction to undertake the research instead, depending on the transferability/generalisability of the results to the local jurisdiction. This could lead to a suboptimal (Nash) equilibrium with a failure to carry out research that would be beneficial to both jurisdictions. Alternatively, there may be a global optimal allocation of patients across jurisdictions in a particular trial, dependent on the relative costs and benefits in each location [32].

In conclusion, VoI is a technique for quantifying the expected return on investment in research. This paper, along with the accompanying Excel files, is intended to provide a useful template that can be readily adapted to other situations.

Appendix 1: Numeric Method to Calculating VoI Statistics

The numeric example presented here is based around a hypothetical (and much simplified) decision model constructed in Microsoft Excel (ESM, Appendix 3). The layout of the spreadsheet model is broadly consistent with that used in examples elsewhere (e.g. [33]). Some familiarity with Excel and macro programming is required to follow the steps described, for which plenty of guides can be found on the World Wide Web.

Suppose a new treatment had been developed for a given disease. Current data suggest it leads to a slightly lower response rate to the disease, but due to its mechanism of action avoids the risk of side effects.

The decision problem is structured as a decision tree (Fig. 5). A patient prescribed 'Old' has a 20 % probability of experiencing side effects and 80 % probability of responding to treatment. A patient prescribed 'New' is not at risk of the side effects, but has only a 75 % probability of responding. A patient responding to treatment is assumed to have a remaining quality-adjusted life expectancy equivalent to eight QALYs. Those that do not respond accrue only four QALYs. Patients experiencing side effects from the Old treatment lose an additional two QALYs. Old costs £500 per patient, but an additional
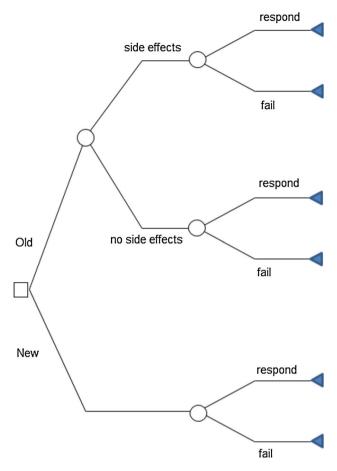
Fig. 5 Decision tree

£1,000 to treat side effects should they occur, whilst -
New costs £2,500.

Input parameters and associated uncertainty are descri-
bed in Table 1 as well as the complete worked example in
the ESM, Appendix 3. Parameter uncertainty is propagated
through the model to characterise decision uncertainty
using Monte Carlo simulation, generating an empirical
distribution of incremental net benefit. Details of how to do
this are available from numerous textbooks (e.g. [24, 34]).

Expected Value of Perfect Information

The EVPI can be expressed as the expected maximum net
benefit with perfect information less the maximum expec-
ted net benefit with current information (Eq. 42).

Example

Table 2 illustrates the method showing the results from just
five simulations (1,000 or more are generally recom-
mended to fully characterise model uncertainty, although
the required number is dependent on a number of factors

Table 1 Input parameters to decision model

Parameter	Distribution	a^a	b^a
Probabilities			
Response on Old	Beta	80	20
Response on New[b]	Beta	75	25
Ln (OR)[b]	Normal	−0.288	0.116
Risk of side effects on Old	Beta	20	80
Costs			
Old	Constant	–	–
New	Constant	–	–
Side effects	Gamma	2.5	400
QALYs			
Respond	Normal	8	0.25
Not respond	Normal	4	0.25
Reduction due to SE	Normal	2	0.04

Ln natural log, *OR* odds ratio, *QALYs* quality-adjusted life-years, -
SE side effects

[a] Parameters for beta distribution are 'successes' and 'failures', for
normal are mean and variance of mean, and for gamma are shape and
scale, respectively

[b] Note, response on New is not entered into the model as an inde-
pendent beta distribution but as the probability of response on old,
converted to odds, multiplied by the odds ratio and then reconverted
back to a probability. The parameters of the distribution of the log
odds ratio are estimated from the 2 × 2 contingency table

including the complexity of the model and level of
uncertainty in input parameters). The net benefit for each
treatment (New and Old) at a threshold of £20,000 is cal-
culated as per Eq. 43 for each of the five simulations. The
final row is the mean (expected) net benefit with each
treatment. In this case, New has the highest (i.e., maxi-
mum) expected net benefit (£142,430 vs £141,125), thus
the decision is to choose New, and 'N' is entered in the
final row of the column 'Decision'. In three of the five
iterations, choosing New would indeed have been correct,
however for iterations 3 and 5, Old yielded a higher net
benefit. There is thus an opportunity loss: in iteration 3, the
maximum net benefit would have been with Old
(£149,245), but as New is chosen, there is a loss of £3,403
(£149,245 − £145,842, shown in column 'Opp. Loss').
Likewise in simulation 5, the opportunity loss would be
£4,837. The expected opportunity loss is therefore £1,647
(final row of column 'Opp. Loss'), which is by definition
the EVPI (per patient). This must be multiplied by the
beneficial population to estimate the overall EVPI to
society.

Worksheet 'PSA' in ESM, Appendix 3 illustrates the
EVPI calculation with 1,000 Monte Carlo simulations. The
macro 'ew_PSA' samples from the input distributions,
recalculates the model and inserts the resulting cost and
QALYs gained from each intervention into columns B to E.

Table 2 EVPI via simulation

Simulation	Net benefit		Decision	Max NB	Opp. Loss
	New	Old			
1	£144,899	£143,643	N	£144,899	£0
2	£141,907	£134,018	N	£141,907	£0
3	£145,842	£149,245	O	£149,245	£3,403
4	£146,496	£140,876	N	£146,496	£0
5	£133,004	£137,841	O	£137,841	£4,837
Expectations	£142,430	£141,125	N	£144,077	£1,647

EVPI Expected value of perfect information, *Max NB* maximum net benefit, *Opp. Loss* opportunity loss, *N* treatment 'New', *O* treatment 'Old'

The net benefit at the threshold set in cell I1 is calculated in columns I and J, with the maximum shown in column M. The expectations are in cells I3, J3 and M3, respectively, with the EVPI calculated in cell N3. (Note, running the macro 'ew_CEAC' is required to update the cost-effectiveness acceptability curve displayed on the worksheet.)

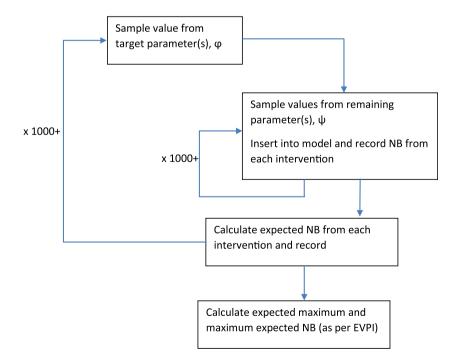### Expected Value of Perfect Parameter Information

The EVPPI is defined in Eq. 44. Note, firstly the similarity with Eq. 42, and secondly the nested expectations: the process to estimate the EVPPI is shown in Fig. 6. The first step is to sample a value from the target parameter or group

of parameters, ϕ. This is one possible realisation of the 'true' value of the parameter(s). A value is then sampled from the remaining parameters, ψ. The parameter set is then inserted into the model and the net benefit from each treatment calculated. A new set of values for ψ is then drawn and, along with the previously drawn values of ϕ, are inserted into the model again and the net benefit recorded. This 'inner loop' is repeated a 'large' number of times (e.g., 1,000 or 5,000), from which the expected net benefit of each treatment is calculated and kept. The outer loop now iterates where a new (set of) value(s) for ϕ is drawn, and the inner loop is repeated. After repeating the outer loop a 'large' number of times, there will be many estimates of the (expected) net benefit from each treatment. Taking the expectation of these, and choosing the maximum is the maximum expected net benefit with current information; i.e., the second term inside the brackets of Eq. 44. The expected maximum net benefit (first term inside the brackets of Eq. 44) is calculated as for the EVPI as the expectation of the maximum net benefit from each iteration.

Example

The summary table for calculating the EVPPI has exactly the same format as for the EVPI (Table 2), where each row represents one iteration of the outer loop, and the numbers recorded are the expected net benefit estimated from the

Fig. 6 Process for EVPPI. ϕ set of target parameter(s) of interest, ψ remaining parameters in decision model. *EVPI* expected value of perfect information, *EVPPI* expected value of perfect parameter information, *NB* net benefit

inner loop. ESM, Appendix 3 illustrates a worked example with 100 outer loops and 1,000 inner, and a macro 'ew_EVPPI' which calculates the EVPPI for three groups of parameters: probabilities, costs and QALYs. The macro first samples a value for the response rate on Old, the odds ratio of response on New and the risk of side effects on Old (sheet 'Inputs', cells J5, J7 and J11). The model is then run for 1,000 iterations holding these values constant whilst values for costs and QALYs are sampled and the results entered into the sheet 'PSA' as before. The expectations, calculated in cells I3 and I4 on sheet 'PSA', are then copied to cells B5 and B6 on sheet 'EVPPI'. As described above, the outer loop then reiterates with a new set of values chosen for the probabilities. After 100 outer loops, the expectation of the expected net benefits for each treatment are calculated in sheet 'EVPPI', cells B3 and B4, with the expected maximum in cell B5. The EVPPI is then calculated as per the EVPI and is in cell D2. When repeated for the three groups of parameters, the results can be shown as a chart (Fig. 7). In this case, the EVPPI is concentrated in uncertainty in probabilities, with very little value to reducing uncertainty in QALYs, and none at all in reducing uncertainty in cost. Again, this per patient EVPPI must be multiplied by the beneficial population to estimate the societal EVPPI.

Expected Value of Sample Information

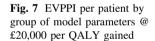The EVSI can be considered as the expected maximum expected net benefit with the new information yielded from a study of sample size n per arm less the maximum expected net benefit with current information, multiplied by the beneficial population less those enrolled in the study (Eq. 45). The second term in the equation is common to Eqs. 42 and 44. The first term is again calculated via simulation with a nested inner and outer loop.
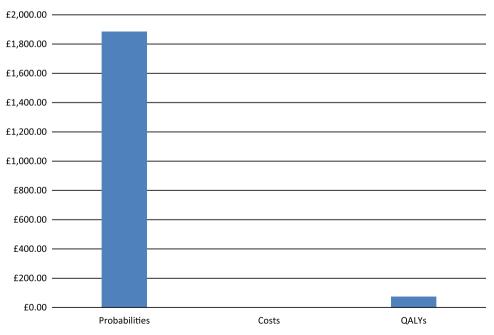
The general approach is to repeatedly predict the results of a trial collecting data on the target parameter(s) based on the prior distributions and incorporating that into a predicted posterior which is then sampled from repeatedly (along with other input parameters in the model, Fig. 8). This entire process must be repeated for a wide range of values of n. The distribution of the sampled data will be related to the prior (a relationship known as conjugacy). A detailed discussion of conjugate distributions may be found by elsewhere [14, 35], but Ades et al. [36] provide a useful set of algorithms for a number of distributional forms.

Example

ESM, Appendix 3 illustrates an example implementing the EVSI for a beta, normal and gamma distribution, as well as methods for calculating the EVSI for an odds ratio (see worksheet 'EVSI').
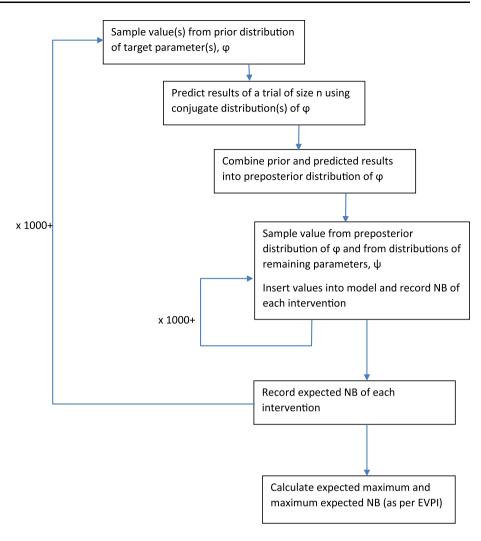
The baseline response rate illustrates the method for calculating EVSI with a beta prior and binomial likelihood. The prior information, based on 100 observations, is in cells B5:B6. These are simply taken from cells G5:H5 on worksheet 'Inputs' and shown in Table 1. A possible value

Fig. 7 EVPPI per patient by group of model parameters @ £20,000 per QALY gained

EVPPI = Expected Value of Perfect Parameter Information; QALY = Quality adjusted life year

Fig. 8 Process for EVSI. ϕ set of target parameter(s) of interest, ψ remaining parameters in decision model

EVSI = Expected value of sample information; φ = set of target parameter(s) of interest; ψ = remaining parameters in decision model; NB= Net Benefit; EVPI = Expected value of perfect information

for the 'true mean' is sampled in cell D5. The macro 'ew_EVSI_BLResp' inserts a proposed sample size for a new study, (ranging between 1 and 2,000). Cell F5 samples a possible trial result from the binomial likelihood using the 'BINOM.INV' function as a possible number of responders out of the total 'n'. The preposterior distribution is defined in cells G5:H5, simply by adding the number of responders to the prior 'a' parameter and non-responders to the 'b' parameter. The macro then inserts this preposterior into cells G5:H5 in worksheet 'Inputs', runs the probabilistic sensitivity analysis (macro 'ew_PSA'), and records the expected net benefit from cells I3:J3 in worksheet 'PSA' in the cells I5:J5 of worksheet 'EVSI'. Cell K5 then chooses the maximum of the two. After this the macro copies and pastes the cells B5:K5 to row 6 before selecting a new possible value for the 'true mean' in cell B5 again and repeating a 'large' number of times (currently set to

500, but in reality many more than this may be required). The EVSI is then calculated in cell K2, based on the summaries calculated in cells I3:K3. The macro copies the EVSI to cell N5.

At this point the entire process is repeated with a new proposed sample size (in the example a study with $n = 10$).

Columns P:AF illustrate the same process but for a normally distributed parameter (here the QALYs gained for a responder). The macro programming is identical (macro 'ew_EVSI_QALYsResp'). The difference is in the calculation of the preposterior distribution. Suppose the prior data are based on a sample size of $n = 100$ (Cell S5). Where the sample size is known this can be entered directly. However, where the sample size is unknown, or based on some structured elicitation exercise (e.g., [37]), a notional sample size can be inferred from the square of the ratio of the standard error and standard deviation as

described in the manuscript for the analytic approach (the standard error being elicited along with the mean, and the standard deviation being estimated from a review of the literature of similar parameters).

The third example estimates the EVSI on a cost study looking at treatment side effects with a gamma distribution. This is most easily handled by sampling from the distribution of the natural log of prior costs as this will be approximately normal, [38, 39] and is indeed what the code shown in cells AH5:AU5 does. The method is then identical to that described for data on QALYs as described above. Where the parameter of interest is count data (e.g., health service contacts), it is possible to program Excel to sample from a Poisson distribution, but the command is not currently inbuilt. However, code is available online to do this (e.g., [40]).

The final example estimates the EVSI of a trial to predict the odds ratio of response to New versus Old. This is done as a two-stage approach [36] whereby firstly the number of responders with baseline treatment (Old) is predicted from the respective prior, then the number of responders in the patients treated with New is predicted by sampling from the log odds ratio. The prior and the data are combined and the resulting preposterior parameters of the log odds ratio are then calculated as per cells CA5:CB5.

Expected Net Gain of Sampling

The approach to calculating ENGS is identical to the analytic method described in the manuscript Sect. 3.4.

$$\text{EVPI} = N\left[E_\theta \max_j \text{NB}(j, \theta) - \max_j E_\theta \text{NB}(j, \theta)\right] \quad (43)$$

where: N = beneficial population (manuscript Eq. 26), Θ = set of input parameters to the decision model, j = intervention/arm, NB_j = net benefit from treatment j, derived from Eqs. 1 and 2 as:

$$\text{NB}_j = \lambda E_j - C_j \quad (44)$$

$$\text{EVPPI}_\varphi = N\left[E_\varphi \max_j E_{\psi|\varphi} \text{NB}(j, \varphi, \psi) - \max_j E_\theta(j, \theta)\right] \quad (45)$$

where: φ is parameter(s) of interest, ψ is other parameters such that $\varphi \cup \psi = \theta$, N, j and NB_j are as per Eq. 27

$$\text{EVSI}_{\varphi,0} = (N - 2n)\left[E_D \max_j E_{\varphi|D} \text{NB}(j, \psi) - \max_j E_\psi(j, \psi, \varphi)\right] \quad (46)$$

References

1. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. In: Senn S, Barnett V, editors. Statistics in practice. Chichester: Wiley; 2004. p. 25.

2. Pratt J, Raiffa H, Schlaifer R. Introduction to statistical decision theory. Cambridge: Massachusetts Institute of Technology; 1995.

3. Henriksson M, Lundgren F, Carlsson P. Informing the efficient use of health care and health care research resources—the case of screening for abdominal aortic aneurysm in Sweden. Health Econ. 2006;15(12):1311–22.

4. Garside R, Pitt M, Somerville M, et al. Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling. Health Technol Assess. 2006;10(8):1–158.

5. Robinson M, Palmer S, Sculpher M, et al. Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling. Health Technol Assess. 2005; 9(27):3–4 (9–11, 1–158).

6. Tappenden P, Chilcott JB, Eggington S, et al. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. Health Technol Assess. 2004;8(27):3 (1–78).

7. Iglesias CP, Claxton K. Comprehensive decision-analytic model and Bayesian value-of-information analysis: pentoxifylline in the treatment of chronic venous leg ulcers. Pharmacoeconomics. 2006;24(5):465–78.

8. Castelnuovo E, Thompson-Coon J, Pitt M, et al. The cost-effectiveness of testing for hepatitis C in former injecting drug users. Health Technol Assess. 2006;10(32):3–4 (9–12, 1–93).

9. Bartell SM, Ponce RA, Takaro TK, et al. Risk estimation and value-of-information analysis for three proposed genetic screening programs for chronic beryllium disease prevention. Risk Anal. 2000;20(1):87–99.

10. Wilson EC, Gurusamy K, Samraj K, et al. A cost utility and value of information analysis of early versus delayed laparoscopic cholecystectomy for acute cholecystitis. Br J Surg. 2010;97:210–9.

11. Gurusamy K, Wilson E, Burroughs AK, et al. Intra-operative vs pre-operative endoscopic sphincterotomy in patients with gallbladder and common bile duct stones: cost-utility and value-of-information analysis. Appl Health Econ Health Policy. 2012;10(1):15–29. doi:10.2165/11594950-000000000-00000.

12. Wilson EC, Emery JD, Kinmonth AL, et al. The cost-effectiveness of a novel SIAscopic diagnostic aid for the management of pigmented skin lesions in primary care: a decision-analytic model. Value Health. 2013;16(2):356–66. doi:10.1016/j.jval.2012.12.008.

13. Raiffa H, Schlaifer R. Probability and statistics for business decisions. New York: McGraw Hill; 1959.

14. Raiffa H, Schlaifer R. Applied statistical decision theory. Boston: Harvard Business School; 1961.

15. Thompson MS. Decision-analytic determination of study size. The case of electronic fetal monitoring. Med Decis Making. 1981;1(2):165–79.

16. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18(3):341–64 (S0167-6296(98)00039-3 [pii]).

17. Eckermann S, Willan AR. Expected value of information and decision making in HTA. Health Econ. 2007;16(2):195–209. doi:10.1002/hec.1161.

18. Willan A, Briggs A. Power and sample size determination: the value of information approach. Statistical analysis of cost-effectiveness data. Chichester: Wiley; 2006. p. 108–16.

19. Banta HD, Thacker SB. The case for reassessment of health care technology: once is not enough. JAMA. 1990;264:235–40.

20. Fenwick E, Claxton K, Sculpher M, et al. Improving the efficiency and relevance of health technology assessment: the role of decision analytic modelling. Centre for Health Economics

Discussion paper 179. York: Centre for Health Economics, University of York; 2000.

21. Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. J Health Serv Res Policy. 1997;2(1):26–30.

22. Sculpher MJ, Claxton K, Drummond M, et al. Whither trial-based economic evaluation for health care decision making? Health Econ. 2006;15(7):677–87.

23. Wilson E, Abrams K. From evidence based economics to economics Based evidence: using systematic review to inform the design of future research. In: Shemilt I, Mugford M, Vale L, et al., editors. Evidence based economics. London: Blackwell Publishing; 2010.

24. Drummond M, Sculpher M, Torrance G, et al. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press; 2005.

25. Cachon GR, Terwiesch C. Matching supply with demand: an introduction to operations management. 3rd ed. New York: McGraw-Hill; 2013.

26. Willan AR, Pinto EM. The value of information and optimal clinical trial design. Stat Med. 2005;24(12):1791–806.

27. Steuten L, van de Wetering G, Groothuis-Oudshoorn K, et al. A systematic and critical review of the evolving methods and applications of value of information in academia and practice. Pharmacoeconomics. 2013;31(1):25–48. doi:10.1007/s40273-012-0008-3.

28. Philips Z, Claxton K, Palmer S. The half-life of truth: what are appropriate time horizons for research decisions? Med Decis Making. 2008;28(3):287–99. doi:10.1177/0272989X07312724.

29. McKenna C, Claxton K. Addressing adoption and research design decisions simultaneously: the role of value of sample information analysis. Med Decis Making. 2011;31(6):853–65. doi:10.1177/0272989X11399921.

30. Eckermann S, Willan AR. The option value of delay in health technology assessment. Med Decis Making. 2008;28(3):300–5. doi:10.1177/0272989X07312477.

31. Bernanke BS. Irreversibility, uncertainty, and cyclical investment. Q J Econ. 1983;98(1):85–106. doi:10.2307/1885568.

32. Eckermann S, Willan AR. Globally optimal trial design for local decision making. Health Econ. 2009;18(2):203–16. doi:10.1002/hec.1353.

33. Briggs AH, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

34. Briggs AH, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Handbooks in health economic evaluation. Oxford: Oxford University Press; 2006. p. 95.

35. Gelman A, Carlin JB, Stern HS, et al. Bayesian data analysis. London: Chapman and Hall; 1995.

36. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. Med Decis Making. 2004;24(2):207–27. doi:10.1177/0272989X04263162.

37. O'Hagan A, Buck CE, Daneshkhah A, et al. Uncertain judgements: eliciting experts' probabilities. Chichester: Wiley; 2006.

38. Higgins JP, White IR, Anzures-Cabrera J. Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. Stat Med. 2008;27(29):6072–92. doi:10.1002/sim.3427.

39. Mathworks Documentation Center: Lognormal mean and variance. http://www.mathworks.co.uk/help/stats/lognstat.html. Last Accessed 11 January 2014.

40. MrExcel.com, Your one stop for Excel tips and solutions. http://www.mrexcel.com/forum/excel-questions/507508-reverse-poisson.html. Last Accessed 11 January 2014.