Efficient Research Design: using Value of Information Analysis to estimate the optimal mix of top-down and bottom-up costing approaches in an economic evaluation alongside a clinical trial

Edward CF Wilson, PhD[1,2]
Miranda Mugford, PhD[2]
Garry Barton, PhD[2]
Lee Shepstone, PhD[3]

1 Cambridge Centre for Health Services Research, Institute of Public Health, University of Cambridge, CB2 0SR

2 Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ, UK

3 Medical Statistics Group, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ.

ed.wilson@medschl.cam.ac.uk

Running title: Efficient choice between data collection methods

Word count:

Acknowledgements

This paper is based on EW's PhD thesis, supervised by the co-authors, LS, MM and GB. The authors wish to thank Dr James Wason, MRC Biostatistics Unit, for comments on previous drafts and two anonymous reviewers for their thoughtful and helpful remarks on this work. Any errors or omissions are the authors' own.

No specific funding was received for this work. EW's thesis was funded by the Norwich Medical School. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Abstract

In designing economic evaluations alongside clinical trials, analysts are frequently faced with alternative methods of collecting the same data, the extremes being top-down ("gross costing") and bottom-up ("micro-costing") approaches.

A priori, bottom-up approaches may be considered superior to top-down but are also more expensive to collect and analyse. In this paper, we use value of information analysis to estimate the efficient mix of observations on each method in a proposed clinical trial.

By assigning a prior bivariate distribution to the two data collection processes, the predicted posterior (i.e. preposterior) mean and variance of the superior process can be calculated from proposed samples using either process. This is then used to calculate the preposterior mean and variance of incremental net benefit and hence the expected net gain of sampling.

We apply this method to a previously collected dataset to estimate the value of conducting a further trial and identifying the optimal mix of observations on drug costs at two 'levels': by individual item ("process A") and by drug class ("process B"). We find that substituting a number of observations on process A for process B leads to a modest £35,000 increase in expected net gain of sampling (ENGS). Drivers of the results are the correlation between the two processes and their relative cost.

This method has potential use following a pilot study, to inform efficient data collection approaches for a subsequent full-scale trial. It provides a formal quantitative approach to inform trialists whether it is efficient to collect resource use data on all patients in a trial, on a subset of patients only, or to collect limited data on most and detailed data on a subset.

1. Introduction

In designing economic evaluations alongside randomised controlled trials, analysts are faced with alternative methods to collect the same data. For example, hospitalisation costs can be estimated using time-and-motion studies and detailed measurement of all drugs dispensed, tests conducted and appropriate allocation of overhead costs; alternatively they can be approximated on a per-admission or per bed-day basis. Drug costs can be estimated by quantifying exact consumption of every drug by every patient; alternatively they can be approximated based on recorded prescriptions of a particular drug or drug class and assumptions over dose and frequency.

The two extremes are known as top-down or gross costing and bottom-up or micro-costing. The choice can, to a certain extent, be determined by the study question. For example, in an economic evaluation comparing two surgical procedures, it would be appropriate to micro-cost the index procedure. However, resources may only permit a more top-down approach to costing other elements such as post-operative length of stay, readmission etc. Indeed, the considerable effort required to accurately measure and value resource use has long been recognised,(1) and some quite dramatic reductions in the cost of projects as a result of careful scrutiny of trial logistics have been documented.(2, 3)

There are numerous examples of comparisons of alternative approaches to collecting the same data (e.g.(4-13)), but very few attempts to quantify the cost-effectiveness of one approach compared with another, and thus to judge when a detailed approach is warranted or whether a more approximate approach is sufficient for purpose, releasing scarce research funds for greater benefit elsewhere. One study that did consider the cost-effectiveness of research focused on a comparison of a prospective RCT versus a retrospective study design,(14) rather than on collecting specific (resource use) elements within a proposed RCT.

In this paper we present an adaptation of the principles of value of information analysis (15-18) to compare the expected return on investment from collecting data using one process compared with another. The scenario we analyse is where a bottom-up data collection process is considered *a priori* superior to a top-down process, but is also more expensive. We apply the method and set it in context with a previously collected dataset by firstly predicting whether a repeat of the trial to further reduce decision uncertainty would be efficient, and if so, the optimal mix of observations on a specific parameter using two data processes (defined as the mix that maximises the expected return on investment).

2. Method

In this section we present a narrative explanation of the principles followed by the algebra. We then describe the analyses to be presented in the results. Appendices detailing all working are provided as online supplementary material. A Microsoft Excel spreadsheet with all calculations is available on request from the corresponding author.

2.1 Narrative explanation

2.1.1 Value of information analysis

Value of information analysis is a technique to predict the expected return on investment in research and is rooted within Bayesian statistical decision theory.(15, 19) The Bayesian statistical approach of defining a prior and updating it with data to generate a posterior is known as posterior analysis. Value of information analysis involves predicting the data based on the prior, which are then combined with the prior to generate a predicted posterior (or 'preposterior') distribution, and so is sometimes known as preposterior analysis.(15)

Suppose a new intervention is proposed to replace an existing treatment. Whether this represents an overall increase in health to an economy is determined by the (mean) incremental net benefit, which we denote ΔB. This is a rearrangement of the incremental cost-effectiveness ratio(20) and if positive, the new intervention should be adopted; if negative, the existing treatment should be retained. Decision makers are assumed risk neutral(21) and thus make adoption decisions on expected values only, irrespective of uncertainty.

Decision uncertainty is represented by the (prior) probability distribution of ΔB, denoted $f(\Delta B)_0$ in Figure 1. The mean is positive so the decision would be to adopt the new treatment. However, the proportion of the probability mass to the left of the Y-axis shows that there is a probability (approximately 33% in this example, the sum of the two shaded areas) that ΔB is negative and a decision to adopt would be wrong. The expected loss due to uncertainty is approximately the probability of being wrong multiplied by the average consequence of being wrong, (18)(22) i.e. the absolute value of the area under the $f(\Delta D)_0$ curve from $-\infty$ to zero. Equivalently, this is the expected gain from eliminating that uncertainty, or the expected value of perfect information (EVPI).

New information (e.g. a clinical trial, database analysis or survey) is expected to reduce the standard error of mean incremental net benefit and so will tighten the distribution around the (updated) mean, yielding $f(\Delta B)_1$ in Figure 1. The proportion of the probability mass to the left of the Y-axis will therefore decrease (dark shaded area) and thus reduce the expected loss (i.e. the predicted posterior EVPI will be less than the prior EVPI). The expected reduction in the expected loss is the expected gain from a particular study, or the expected value of sample information (EVSI). A larger study will provide more information than a smaller one, but also cost more. The EVSI of a study of sample size n less the (expected) cost is the expected net gain of sampling (ENGS). The predicted optimal sample size for a new study, n*, is that which maximises ENGS.

It may be more efficient to concentrate data collection on one or more components of ΔB, such as health outcomes or some component of cost. To calculate the ENGS of such a study, the expected reduction in standard error of that component from n observations is followed through to the expected reduction in standard error of ΔB (i.e. the expected reduction in parameter uncertainty is translated into an expected reduction in decision uncertainty).(18) The EVSI and ENGS are then calculated as before.

2.1.2 Extension of principles to compare alternative data collection processes

Pratt, Raiffa & Schlaifer provide an extension to these principles to compare two alternative data collection processes.(15) We adapt their technique to the healthcare field as follows:

Given a choice between a top-down and bottom-up approach to calculating a component of ΔB (for example incremental cost of drugs), the prior distribution of the two is assumed bivariate Normal. Note the covariance provides information on the relationship between the two. Such data could be obtained from a pilot study where both approaches (hereafter termed 'data processes') are observed in the same patient group, a review of the literature, or elicited from experts (e.g. (23, 24)).

As stated, we assume the bottom-up process is superior to the top-down in that it is a more accurate measure of cost (that is, it provides the least biased estimate of the mean and the most appropriate characterisation of the dispersion of individual costs around the mean).

We label the bottom-up process A and the top-down process B. The estimate of (mean incremental) drug cost yielded from process A, ΔC_d^A should be used in the calculation of ΔB as it is believed to be a 'better' estimate than that yielded from process B, ΔC_d^B (the subscript 'd' refers to drugs). Specifying a prior bivariate distribution allows one to determine how belief about ΔC_d^A should be revised given information on ΔC_d^B alone, or a mix of information on ΔC_d^A and ΔC_d^B. In other words, given a reduction in standard error of ΔC_d^B from n observations on ΔC_d^B, it is possible to predict the expected reduction in standard error of ΔC_d^A, which is then followed through to a predicted reduction in standard error of ΔB (i.e. reduction in decision uncertainty, Figure 2). The EVSI and the ENGS of the proposed study can then be calculated. This approach is repeated with combinations of sample sizes for observations on ΔC_d^A and ΔC_d^B. The ENGS-maximising combination is the optimal combination. (Note that we assume that only one process is observed in each individual. This is a limitation of the method and is considered in the discussion.)

2.2 Algebraic explanation

This explanation comprises three sections. In the first, the basic model is set up linking prior data or expert beliefs explicitly to distributions of incremental net benefit and its components. The second section explains the relationship between the two data collection processes for the incremental cost of drugs. The final section briefly explains how the value of information statistics are calculated; more detailed explanations of these are available elsewhere e.g. (18, 25).

2.2.1 Basic model: means, variance and covariance

The objective is to maximise expected net benefit, which can be expressed as choosing the option with the highest expected net benefit, or where there are only two treatment options, choosing new treatment (T) in place of current practice (C) if the incremental net benefit of T compared with C is positive. Define mean net benefit per patient in treatment arm j, B_j, as the value of mean health gain (QALYs gained, E, multiplied by the value attached to a QALY, λ), less the mean cost (equation [1]).

$$B_j = \lambda E_j - C_j \qquad j = T, C\ (Treatment\ and\ Control\ respectively) \tag{1}$$

Here, cost comprises just two components: cost of drugs, C_d^A, and all other (non-drug) costs, C_n (equation [2]). (The superscript 'A' is explained below).

$$C_j = C_{n,j} + C_{d,j}^A \qquad j = T, C \tag{2}$$

Where individual patient data are available, mean costs and QALYs can be calculated directly (equation [3]). Alternatively they may be based on a meta-analysis of existing data or expert beliefs (e.g. (23, 24)).

$$X_j = \frac{\sum_{i=1}^{n_j} x_{i,j}}{n_j} \qquad j = T, C;\ \ X = E, C_n, C_d^A;\ \ x_i = e_i, c_{n,i}, c_{d,i}^A;\ \ i = patient; \tag{3}$$

E_j = mean QALYs in arm j,

$C_{n,j}$ = mean non − drug costs in arm j,

$C^A_{d,j}$ = mean drug costs (using process A)in arm j;

$e_{i,j}$ = QALYs gained by patient i in arm j,

$c_{n,i,j}$ = non − drug costs in patient i, arm j,

$c^A_{d,i,j}$ = drug costs (using process A) in patient I, arm j.

(Mean) incremental net benefit, ΔB, can be defined as the difference in (mean) net benefit between each course of action (B_T and B_C respectively; equation [4]). Note that Equation [4] can also be derived from a rearrangement of the incremental cost effectiveness ratio (ICER).

$$\Delta B = B_T - B_C \quad T = treatment, C = control \tag{4}$$

The variance of ΔB, v(ΔB) is therefore the sum of the variances of net benefit in each arm (equation [5]).

$$v(\Delta B) = v(B_T) + v(B_C) \tag{5}$$

As net benefit in each arm is a linear function of cost and outcome, and cost is a linear function of drug and non-drug costs, the variances of each are as per equations [6] and [7].

$$v(B_j) = \lambda^2 v(E_j) + v(C_j) - 2\lambda Cov(E_j, C_j) \qquad j = T, C \tag{6}$$

$$v(C_j) = v(C^A_{d,j}) + v(C_{n,j}) - 2\lambda Cov(C^A_{d,j}, C_{n,j}) \qquad j = T, C \tag{7}$$

As before, the variances and covariances can be calculated from trial data (Equations [8-9]) or estimated from meta-analyses and/or expert opinion. We adopt the convention of a lower case letter denoting an individual observation whilst uppercase denotes the population mean. Thus $e_{i,j}$ is the QALYs gained by patient i in arm j, whilst E_j is the mean QALYs gained per patient in arm j. As such $v(e_j)$ is the sample variance of QALYs in arm j, whilst $v(E_j)$ is the variance of the mean (Equation [8]). It is of critical importance not to confuse these two, or their square roots (the standard deviation and standard error of the mean respectively).

$$v(X_j) = \frac{1}{n_j} . \frac{\sum_{i=1}^{n_j} (x_{i,j} - X_j)^2}{(n_j - 1)} \qquad j = T, C; \quad X = E, C_n, C^A_d; \; x_{i,j} = e_{i,j}, c_{n,i,j}, c^A_{d,i,j}; \tag{8}$$

$i = individual; j = arm$

$$Cov(X_j, Y_j) = \frac{1}{n_j} . \frac{\sum_{i=1}^{n_j} (x_{i,j} - X_j)(y_{i,j} - Y_j)}{(n_j - 1)} \qquad j = T, C; \tag{9}$$

$$\{X, Y\} = \{C_n, C_d^A\}, \{E, C\}$$

Inserting equation [6] into [5] provides an alternative expression for v(ΔB) as the sum of the variances of incremental cost and outcomes less twice the respective covariances (equation [10]).

$$v(\Delta B) = \lambda^2 v(E_T) + v(C_T) - 2\lambda Cov(E_T, C_T) + \lambda^2 v(E_C) + v(C_C) - 2\lambda Cov(E_C, C_C) \quad [10]$$
$$= \lambda^2\big(v(E_T) + v(E_C)\big) + v(C_T) + v(C_C) - 2\lambda\big(Cov(E_T, C_T) + Cov(E_C, C_C)\big)$$
$$= \lambda^2 v(\Delta E) + v(\Delta C) - 2\lambda Cov(\Delta E, \Delta C)$$

Expressing the covariance as the product of the correlation coefficient (ρ) and the standard errors (equation [11]) and with the subscript '0' denoting the priors yields equations for the prior variance of incremental net benefit as a whole (equation [12]) and incremental cost specifically (equation [13]).

$$Cov(X, Y) = \rho_{X,Y}\sqrt{v(X)}\sqrt{v(Y)} \quad [11]$$

$$v(\Delta B)_0 = \lambda^2 v(\Delta E)_0 + v(\Delta C)_0 - 2\lambda \rho_{\Delta E, \Delta C, 0}\sqrt{v(\Delta E)_0}\sqrt{v(\Delta C)_0} \quad [12]$$

$$v(\Delta C)_0 = \lambda^2 v(\Delta C_n)_0 + v\big(\Delta C_d^A\big)_0 + 2\lambda \rho_{\Delta C_n, \Delta C_d^A, 0}\sqrt{v(\Delta C_n)_0}\sqrt{v(\Delta C_d^A)_0} \quad [13]$$

Note that there are five parameters to $v(\Delta B)_0$ (equations [12-13]): not only $v(\Delta E)$, $v(\Delta C_n)$, and $v\big(\Delta C_d^A\big)$, but also $\rho_{\Delta E, \Delta C}$ and $\rho_{\Delta C_n, \Delta C_d^A}$, information on any of which could be used to revise the variance of ΔB to its posterior, $v(\Delta B)_1$.

2.2.2 Defining the relationship between the alternative data collection processes and calculation of predicted posteriors (preposteriors) following proposed data collection.

Now assume that process B, qualitatively inferior to A, is available to estimate the incremental cost of drugs. Call this ΔC_d^B. Given prior belief that A is 'superior', ΔC_d^A should be used in the calculation of ΔB. However, knowledge of the relationship between ΔC_d^A and ΔC_d^B allows revision of beliefs about ΔC_d^A in the light of information on ΔC_d^B. The logic is as follows:

The prior expectations and variance/covariance matrix (of the means) are in Equation [14].

$$E\begin{bmatrix}\Delta C_d^A \\ \Delta C_d^B\end{bmatrix} = \begin{bmatrix}\Delta C_{d,0}^A \\ \Delta C_{d,0}^B\end{bmatrix}, \qquad V\begin{bmatrix}\Delta C_d^A \\ \Delta C_d^B\end{bmatrix} = \begin{bmatrix} V\left(\Delta C_d^A\right)_0 & Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ Cov(\Delta C_d^A, \Delta C_d^B)_0 & V\left(\Delta C_d^B\right)_0 \end{bmatrix} \quad [14]$$

Suppose some data were to be collected on ΔC_d^A and ΔC_d^B. The sample means, denoted $\Delta C_{d,s}^A$ and $\Delta C_{d,s}^B$ with sample sizes n_A and n_B respectively, have expectations and variances/covariances as per Equation [15], where $v(\Delta c_d^A)_s/n_A$ and $v(\Delta c_d^B)_s/n_B$ are the variances of the means estimated from the sample data according to equation [8] (note the lower case ‘c’ denoting the sample variance).

$$E\left(\begin{bmatrix}\Delta C_{d,s}^A \\ \Delta C_{d,s}^B\end{bmatrix} \middle| \begin{bmatrix}\Delta C_d^A \\ \Delta C_d^B\end{bmatrix}\right) = \begin{bmatrix}\Delta C_{d,0}^A \\ \Delta C_{d,0}^B\end{bmatrix}, \quad [15]$$
$$V\left(\begin{bmatrix}\Delta C_{d,s}^A \\ \Delta C_{d,s}^B\end{bmatrix} \middle| \begin{bmatrix}\Delta C_{d,0}^A \\ \Delta C_{d,0}^B\end{bmatrix}\right) = \begin{bmatrix} v(\Delta c_d^A)_s/n_A & 0 \\ 0 & v(\Delta c_d^B)_s/n_B \end{bmatrix}$$

The objective is to combine [15] with [14] to estimate the preposterior distributions. Given the bivariate Normal distribution, this is achieved as follows (notation adapted from Pratt, Raiffa & Schlaifer(15)):

1. Define **H'** as the inverse of the prior var/covar matrix (Equation [16])

$$\boldsymbol{H'} = \begin{bmatrix} H'_{11} & H'_{12} \\ H'_{21} & H'_{22} \end{bmatrix} = \begin{bmatrix} v\left(\Delta C_d^A\right)_0 & Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ Cov(\Delta C_d^A, \Delta C_d^B)_0 & v\left(\Delta C_d^B\right)_0 \end{bmatrix}^{-1} \quad [16]$$
$$= \frac{1}{v(\Delta C_d^A)_0 v(\Delta C_d^B)_0 - Cov(\Delta C_d^A, \Delta C_d^B)_0{}^2} \begin{bmatrix} v\left(\Delta C_d^B\right)_0 & -Cov(\Delta C_d^A, \Delta C_d^B)_0 \\ -Cov(\Delta C_d^A, \Delta C_d^B)_0 & v\left(\Delta C_d^A\right)_0 \end{bmatrix}$$

2. Define **H** as the matrix of 1 over each component of Equation [15] (i.e. the precision matrix, Equation [17])

$$\boldsymbol{H} = \begin{bmatrix} {}^{n_A}\!/_{v(\Delta c_d^A)_s} & 0 \\ 0 & {}^{n_B}\!/_{v(\Delta c_d^B)_s} \end{bmatrix} \quad [17]$$

3. Define **H''** as the sum of **H'** and **H** (Equation [18])

$$\boldsymbol{H''} = \boldsymbol{H'} + \boldsymbol{H} = \begin{bmatrix} H'_{11} + {}^{n_A}\!/_{v(\Delta c_d^A)_s} & H'_{12} \\ H'_{21} & H'_{22} + {}^{n_B}\!/_{v(\Delta c_d^B)_s} \end{bmatrix} \quad [18]$$

The posterior variance/covariance matrix is the inverse of **H''.** The posterior distribution is summarised in Equations [19] and [20], where **m** is the matrix of sample means from each data process (Equation [21]).

$$\begin{bmatrix}\Delta C_{d,1}^{A}\\ \Delta C_{d,1}^{B}\end{bmatrix} = \boldsymbol{H}''^{-1}\left(\boldsymbol{H}'\begin{bmatrix}\Delta C_{d,0}^{A}\\ \Delta C_{d,0}^{B}\end{bmatrix} + \boldsymbol{Hm}\right) \tag{19}$$

$$\boldsymbol{V}'' = \begin{bmatrix} v\left(\Delta C_{d}^{A}\right)_1 & Cov(\Delta C_{d}^{A}, \Delta C_{d}^{B})_1 \\ Cov(\Delta C_{d}^{A}, \Delta C_{d}^{B})_1 & v\left(\Delta C_{d}^{B}\right)_1 \end{bmatrix} = \boldsymbol{H}''^{-1} \tag{20}$$

$$\boldsymbol{m} = \begin{bmatrix}\Delta C_{d,s}^{A}\\ \Delta C_{d,s}^{B}\end{bmatrix} \tag{21}$$

Equations [16-21] thus show how the preposterior mean and variance of the bivariate Normal parameters are calculated after proposed collection of (n_A, n_B) observations on each process respectively.

2.2.3 Value of information statistics

The predicted posterior mean and variance of process A ($\Delta C_{d,1}^{A}$ in equation [19] and $v\left(\Delta C_{d}^{A}\right)_1$ of equation [20] respectively) are used to calculate the predicted posterior mean and variance of ΔB (equations [12] and [13]), and thence the ENGS, defined as the EVSI less the cost of sampling (Equation [22]). The EVSI is calculated using the unit Normal linear loss integral (UNLLI, Equations [23] to [25]). The UNLLI is explained in greater detail elsewhere, (18, 25) but briefly, this simply calculates the difference in expected loss between the prior and predicted posterior distributions of ΔB and can be used where ΔB is normally distributed and loss is linear in ΔB.

The total cost of sampling is conventionally simplified to a variable per-patient cost for each data process, $k_{s,A}$ and $k_{s,B}$ respectively, plus a fixed cost, K_s (incurred if either n_A or n_B are greater than zero), plus the expected opportunity loss of patients enrolled into the inferior arm of the study (Equation 26). Calculating for a range of values of n_A and n_B, identifies the combination yielding the highest ENGS.

$$ENGS_{n_A,n_B} = EVSI_{n_A,n_B} - TC_{n_A,n_B} \quad [22]$$

$$EVSI_{n_A,n_B} = \left(N - 2(n_A + n_B)\right).\sqrt{v(\Delta B)_{s,n}}.L_{N^*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) \quad [23]$$

$$\begin{aligned} L_{N^*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) \\ &= \phi\left(\frac{|\Delta B_0|}{\sqrt{v(\Delta B)_{s,n}}}\right) \\ &- \frac{|\Delta B_0|}{\sqrt{v(\Delta B)_{s,n}}}\left[\Phi\left(-\frac{|\Delta B_0|}{\sqrt{v(\Delta B)_{s,n}}}\right) - I\{\Delta B_0 < 0\}\right] \end{aligned} \quad [24]$$

ϕ(x)= standard Normal pdf evaluated at x,

Φ(x) = standard Normal cdf evaluated at x.

$$\begin{aligned} v(\Delta B)_{s,n} &= v(\Delta B)_0 - v(\Delta B)_1 \\ &= v(\Delta B)_0 - \frac{v(\Delta b)}{n_0 + (n_A + n_B)} \\ &= v(\Delta B)_0 - \left(\frac{v(\Delta b)}{\frac{v(\Delta b)}{v(\Delta B)_0} + (n_A + n_B)}\right) \\ &= v(\Delta B)_0 - \left(\frac{1}{v(\Delta B)_0} + \frac{(n_A + n_B)}{v(\Delta b)}\right)^{-1} \end{aligned} \quad [25]$$

$$TC = [k_{sA}n_A + k_{sB}n_B + K_s I\{n_A > 0 \cup n_B > 0\} + (n_A + n_B)|\Delta B_0|] \quad [26]$$

2.3 Analyses and layout of results

The ultimate objective is to identify the optimal mix of top-down and bottom-up observations to collect drug costs, that is, the mix that maximises the expected return on investment, as part of any future study. However we set this in context by also presenting standard value of information analyses on other components of ΔB. This is then broadened to identify the overall optimal number of observations on each drug cost process, as well as other parameters (non-drug costs and QALYs gained), thus providing a decision analytic approach to overall trial design. Therefore, we report the following:

i. Value of information analysis for a repeat of the subject trial.

Analysis of uncertainty in ΔB and standard value of information analysis (reporting the EVPI, EVSI and optimal sample size of a trial reporting ΔB and collecting all data on all patients).

ii. Value of information analysis for studies collecting one component of data alone.

We report analyses pertaining to studies collecting (a) incremental QALYs and (b) incremental cost alone. We then sub-divide cost into two individual studies collecting (c) incremental non-drug costs (ΔC_n) and (d) incremental cost of drugs (ΔC_d^A) alone.

iii. Comparison of value of alternative data collection processes on drug costs.

This is the key analysis of the paper. Here we introduce the top-down 'process B' for collecting drug costs and report the efficient mix of observations between the two measures of drug costs (i.e. n_A and n_B observations on ΔC_d^A and ΔC_d^B respectively.).

iv. Overall efficient design of a future trial

The efficient numbers of observations on ΔC_d^A, ΔC_d^B, non-drug costs (ΔC_n) and QALYs (ΔE) are determined simultaneously in this analysis using a Nelder-Mead search algorithm,(26) providing an overall efficient 'portfolio study'.(27)

2.4 Data

Full details are provided in Appendix 1. Briefly, data are taken from a study of 359 patients randomised to leukotriene receptor antagonists versus conventional treatment in asthma patients in a UK setting (the 'ELEVATE' study).(28, 29) The data were divided into drug and NHS-non drug costs and QALYs gained at two years. For illustrative purposes, ΔB was calculated at a threshold of £5000 per QALY.

Drug costs were originally calculated in a bottom-up manner at individual preparation level, based on actual quantities of drugs prescribed. To simulate a 'top down' approach, the cost for each datum was recalculated at the BNF chapter section level, using aggregate cost per prescription as reported in the Prescription Cost Analysis 2005.(30) Therefore every patient had two estimates of drug costs, one based on actual prescribed doses of drugs and the other an approximation aggregated at the BNF section level. We define process A as the drug costs estimated using actual prescribed doses, and process B as the approximation aggregated at BNF section level.

The resulting summary statistics are in Tables 1-2. At a willingness to pay for a QALY of £5000 and using process A for drug costs, incremental net benefit is £56.41. The adoption decision would therefore be in favour of intervention.

As process 'A' is considered superior to 'B', estimates of mean ΔB are based on data from process A. Nevertheless, for the purpose of illustration, recalculating the results using process B yields an incremental net benefit of -£130.86. Using these data the adoption decision would be in favour of control (Table 1).

The population who could benefit from the information is 524,380 (Appendix 1). The fixed cost of sampling is £1,305,470, with a variable cost of £288.58 per patient to collect all data components (QALYs, non-drug and drug costs, Appendix 1). We assume a per patient variable cost of £192.19 for a trial collecting solely QALY or cost data (2/3^{rds} the full cost), £96.19 for one collecting data on either non-drug cost or drug cost data alone (1/3rd) and £9.62 for one collecting only drug cost data using process B (1/10th the cost of process A). These costs are assumptions based on the authors' opinions and experience as to the relative research effort required to collect and analyse the data.

3. Results

Full details for all analyses are in Appendix 2, boxes A2.1-12, and summarised in Table 3. Additional figures are in Appendix 3.

3.1 Value of information analysis for a repeated trial

As reported in the description of the data, prior mean incremental net benefit is £56.41 with a standard error of £217.15 (Table 1 and 2, Figure A3.1a). The population EVPI is £32.2m (Table 3, Box A2.1). The ENGS-maximising sample size is a trial enrolling 2,277 patients per arm yielding an ENGS of approximately £27.3m (Box A2.2, Figure A3.1b & Table 3). Thus the efficient sample size of a repeat of the trial reporting incremental net benefit as its outcome is 2,277 per arm.

3.2. Value of information analysis of four separate studies reporting incremental QALYs, incremental cost, incremental non-drug cost and incremental drug cost alone

The expected value of eliminating uncertainty in outcomes (QALYs) alone (i.e. $EVPPI_{QALYs}$) is £29.2m, and in cost, £6.759m (Boxes A2.3, A2.5 respectively and Table 3). Optimal sample sizes of trials just collecting QALYs or Cost are 2,473 and 1,585 per arm respectively (Box A2.4, Figure A3.2a & Box A2.6, Figure A3.2b respectively & Table 3).

Further dividing costs into non-drug and drug costs, the EVPPI is £3.943m and £3.433m respectively (Boxes A2.7 and A2.9, Table 3), with optimal sample sizes of studies collecting data on those components alone of 1,947 and 1,852 per arm (Box A2.8 & Figure A3.2c and Box A2.10 & Figure A3.2d, Table 3). Figure A3.3 summarises the EVPI, and EVPPI on QALYs, non-drug costs and drug costs.

3.3 Comparison of value of alternative data collection processes on drug costs

The optimal sample size of a study collecting drug cost data alone is estimated at 1621 observations on ΔC_d^A plus 819 observations on ΔC_d^B per arm (Box A2.11, Table 3). Figure 3 shows a three dimensional plot of the ENGS as a function of the sample size of each component. This peaks at (1621, 819) with an expected net gain of sampling of £0.855m. This compares with £0.820m for a trial collecting data only on ΔC_d^A.

3.4 Overall efficient trial design

Calculating for different combinations of $n_{\Delta E}$, $n_{\Delta Cn}$, $n_{\Delta CAd}$ and $n_{\Delta CBd}$ (that is, the number of observations per arm collecting QALYs, non-drug costs, drug costs using process A and drug costs using process B respectively), the ENGS maximising combination can be identified. The combination is (2913, 1064, 736, 901) for $(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A}, n_{\Delta C_d^B})$, yielding an ENGS of

£27.846m (Box A2.12, Table 3 final row). This compares with the maximum ENGS of a trial reporting INB alone of £27.312m (Table 3, first row).

4. Discussion

4.1 Implications of results

In this paper we demonstrate how value of information analysis can be extended to consider the efficient choice between two methods for collecting the same data, thus providing guidance for researchers planning an economic evaluation alongside a clinical trial.

The results demonstrate high expected value from eliminating all known decision uncertainty (EVPI £32.1m, Table 3), due to both high per patient decision uncertainty (coefficient of variation of 217.15/56.41 = 3.8) and the large population who could benefit from this information. There are very few other VoI studies in respiratory disease. The only other study we identified was of pharmacogenomic approaches to diagnosing non-small-cell lung cancer, which estimated an EVPI to the US economy of $31.4m.(31)

If a trial were proposed with the objective of estimating ΔB, the optimal sample size would be 2,277 per arm, costing £2.6m (plus an opportunity loss of £0.1m leading to a total cost of £2.7m), but would yield an expected net gain of sampling of £27.3m. This would be a large trial, approximately 12 times the size of the original.(28) Nevertheless it is the predicted optimal sample size taking into account the cost of acquiring the data and the expected value of the information to the population.

The key analysis in this paper estimated the expected return on a study of incremental drug costs alone, comparing two alternative approaches to collecting the data. We estimate an optimal mix of 1621 observations using the bottom-up process (A) and 819 observations with the top-down (B). The cost of such a study would be £1.7m (plus £0.1m opportunity loss), yielding an ENGS of £0.855m. By using a mix of both processes, a small increase (of £35,000) in the expected return can be obtained compared with using process A alone (rows 5 and 6 of Table 3).

Finally, the optimal numbers of observations for each data component within one study, including the optimal mix between the two data processes for drug cost data, are 2913 on QALYs, 1064 on non-drug cost, 736 on drug cost using the process A and 901 observations using the process B (row 7, Table 3). This would yield an ENGS of £27.846m, an increase of £534,000 on a trial collecting all data on 2,277 observations. Thus selectivity in data collection in this case leads to a higher expected return on investment. It should also be noted that gathering information on all parameters simultaneously changes the optimal mix of observations on data processes A and B: where only A & B are collected (row 6, Table 3), approximately 2/3^{rds} of observations should be on the superior process A. When other data are also collected, only 45% of the observations should be on process A (736 of 1637, row 7, Table 3).

4.2 Practical Implementation

Our analyses provide an estimate of efficient sample size unconstrained by budget, defining the optimum at the point where the marginal gain from an observation is equal to the marginal cost (analogous to the profit maximising condition in the theory of the firm). When designing a trial, this technique can be used to provide a rational quantitative approach to determining what data to collect on which patients. The prior distributions required for each of the inputs and research cost estimates would ideally be provided by a pilot or feasibility study conducted prior to a full scale trial. Alternatively uncertainty in parameters can be captured via a formal elicitation process.(24)

However, trialists are usually faced with exogenous budget or sample size constraints. To incorporate these constraints, it is simply a question of defining a feasible set of observations on each component such that the cost of sampling is less or equal to the budget, K (Equation 27). In this case, for a maximum budget of £2m the optimal combination of observations on ($n_{\Delta E}$, $n_{\Delta Cn}$, $n_{\Delta CAd}$, $n_{\Delta CBd}$) is (2379, 722, 419, 900). This trial would cost £1,999,981 and yield an ENGS of £27.754m. This solution was identified using the Nelder-Mead algorithm.

Likewise, it is straightforward to choose the optimal mix from a feasible set where the sample size has already been determined (for example through a conventional power calculation based on a clinically important difference in a primary outcome).

$$k_{sE}n_E + k_{sC_n}n_{C_n} + k_{sC_d^A}n_{C_d^A} + k_{sC_d^B}n_{C_d^B} + K_s I\left\{n_E > 0 \cup n_{C_n} > 0 \cup n_{C_d^A} > 0 \cup n_{C_d^B} > 0\right\} \le K \quad [27]$$

Where k_{sX} = variable cost of sampling parameter X; K_s = fixed cost of sampling; I{} = indicator function returning 1 if the expression in parentheses is true.

In this analysis, we focused on a very narrow example, driven by the data available to us. However, there is no reason that the principles cannot be extended to other related problems such as the decision to use routine or administrative data sources in place of study specific data. All that is required is a prior belief that one method is measuring the true quantity of interest, that the other is an approximation, and that there is some prior belief about the relationship between the two.

It should be noted that the class of decision problem where this analysis applies is where there is a genuine choice between two alternative methods to collect the same information, such as top-down versus bottom-up costing, medical records versus patient questionnaires or routine administrative data versus study specific collection. Decisions over collection of different components of cost, such as whether to include dispensing fees in drug cost calculations, or social services costs as well as hospital costs can be analysed using standard value of information approaches.

4.3 Determinants of optimal mix of observations between two processes

The optimal mix of observations on two data processes is a function of the relationship between the two (as expressed in the correlation coefficient, ρ) and the relative cost of sampling. Where the data processes are very closely related (ρ close to ±1), then one would expect the top-down (process B) to be the optimal choice due to the lower cost of sampling: observations on B can be used to revise beliefs about A simply by adjusting for the prior estimate of bias. However, where the processes are less closely related there is a trade-off between the extra cost of A and the extra information it yields compared to B. Where the correlation is zero, gathering information using B provides no information on A, therefore it would never be efficient to use process B.

In the dataset used in this paper, the (prior) correlation coefficient between ΔC_d^A and ΔC_d^B is 0.83. Given this, and the relative cost of processes A & B, it is efficient for 34% (819/2440) of observations to be on process B. It is worth investigating how the optimal mix changes with different values of ρ (Figure 4). As predicted, at almost perfect positive or negative correlation, process B provides equivalent information to process A; as process B is cheaper than process A, it is always preferable to draw observations on that process. As ρ falls, for a given sample size process B provides less information on ΔC_d^A, until the value of the information falls below the marginal cost of sampling at which point it is only worth collecting data using process A.

4.4 Comparison with other studies

The origins of this analysis lie in statistical decision theory, developed in the 1960s at the Harvard Business School.(15) However, we are aware of only one previous application of value of information principles to help choose study designs. Shavit and colleagues(14) presented a method to compare the 'net information benefit' of an RCT with an observational study. They define this as a function of the current evidence and estimates of the magnitude of five discrete sources of bias associated with the two designs: representation, selection, time frame, real-life reflection and accuracy of records. Measures of each bias were quantified and expressed as percentage deviation from the true mean.

A key difference between their approach and ours is the question being asked: Shavit and colleagues(14) are concerned with the choice between a prospective RCT and retrospective cohort study to answer a decision question. Our analysis starts where an RCT design has already been chosen, but the approach to collecting various components of the data is undecided.

4.5 Strengths and Limitations

We present a quantitative method to assist the design of a clinical trial, specifically predicting the numbers of observations on different input parameters to incremental net benefit based on maximising the expected return on investment in research. To this end, there are a number of limitations and assumptions which must be considered.

Firstly, we referred to costing a particular resource item as a 'data process' without explicitly differentiating between measurement and valuation. However, this has no consequence for the analysis: two data collection processes may vary in how resource use data are collected (e.g. medical records vs patient self-report) or by valuation technique (costing individual items vs average unit costs to classes of items). The example presented here simulates a hybrid of the two: the simpler data process collects data at a more aggregate level and applies an average unit cost by drug class based on a representative daily dose.

Secondly, we expressed the covariance between parameters as the product of the correlation coefficient and standard errors. This allows the correlation coefficient to be treated as independent from the variances, and potentially as a parameter about which information could be sought. However, for simplicity, we assumed the correlation coefficient, ρ, to be constant. This raises two issues. Firstly, this only allows a very simple linear relationship between the two parameters and secondly we ignored uncertainty in ρ. The first issue can be handled either by transforming one of the parameters (for example, the relationship may be log-linear), or formal modelling of the relationship between the parameters. The second issue raises additional complications as ρ has a non-normal distribution.

Indeed a key limitation of the analysis is the assumption of Normality: costs are known to be right skewed, whilst QALYs may be left skewed (depending on the patient population). The major advantage of this assumption is ease of computation and availability of analytic solutions, but at the risk of misleading conclusions should a Normal distribution be a poor representation of prior beliefs. Simulation approaches may overcome this, but are computationally expensive. However, techniques have been developed to 'short cut' such processes.(32, 33) The Central Limit Theorem states that the sampling distribution of the mean will be approximately Normal. However this is only true in data with a small coefficient of variation and/or large sample size.(34) Therefore an obvious extension to this work is to consider alternative distributional forms for cost data such as bivariate gamma, or simulation approaches using appropriate software(35) with appropriate programming expertise.

Thirdly, we assumed that the overall sample size in analysis 3.4 would be the maximum of each individual parameter, namely 2913 patients in each arm (on which QALY data would be obtained). Of those, 1064 would be chosen at random from which non-drug cost data would be obtained, then 1637 would be chosen from which drug cost data would be obtained, 736 of which using process A and 901 using process B. The cost of the trial was estimated on this basis. However, where patients provide data on more than one component, the covariance and hence correlation between those components can be estimated and used to revise the prior estimates of the correlation coefficient. Our analysis currently ignores this additional information and so may be overestimating optimal trial sizes. However, incorporating this is not straightforward and is an area for further research.

Other limitations are exclusion of other developments in the application of value of information analysis to the healthcare field, such as the appropriate time horizon for an analysis,(36) delays whilst research is conducted(37) and optimal allocation of projects across different jurisdictions(38) potentially affecting the ENGS of a study. We also assumed a constant marginal cost of recruitment. This is an oversimplification of the cost function as the first patients are likely to be easier to recruit than the last ones, as stocks of 'willing volunteers' get exhausted, and further effort is required to identify new patients.

5. Conclusion

In this paper we have shown how it is possible to adapt the principles of value of information analysis to estimate the optimal mix of bottom-up and top-down data collection processes for a component of resource use data in an economic evaluation alongside a clinical trial. Furthermore we have shown how this can be incorporated within a broader decision analytic approach to estimating efficient sample sizes of different data components. Selectivity in the numbers of observations for each component can help contain cost and yield a higher expected net benefit of sampling than one measuring all data on all patients. Whilst the method presented can be used to help researchers design trials, future work will address the current limitations and incorporate other recent advances in VoI methodology.

7. References

1. Drummond M, Davis L. Economic analysis alongside clinical trials: revisiting methodological issues. Int J Technol Assess Health Care. 1991;7(4):561-73.
2. Urban N, Self S, Kessler L, Prentice R, Henderson M, Iverson D, et al. Analysis of the costs of a large prevention trial. Controlled Clinical Trials. 1990 Apr;11(2):129-46. PubMed PMID: WOS:A1990DB35500006.
3. Thornquist MD, Urban N, Tseng A, Edelstein C, Lund B, Omenn GS. Research cost analyses to aid in decison-making in the conduct of a large prevention trial, CARET. Controlled Clinical Trials. 1993 Aug;14(4):325-39. PubMed PMID: WOS:A1993LQ02000007.
4. Merlo J, Berglund G, Wirfalt E, Gullberg B, Hedblad B, Manjer J, et al. Self-administered questionnaire compared with a personal diary for assessment of current use of hormone therapy: An analysis of 16,060 women. American Journal of Epidemiology. 2000 Oct;152(8):788-92. PubMed PMID: WOS:000089906500013.
5. Patel A, Rendu A, Moran P, Leese M, Mann A, Knapp M. A comparison of two methods of collecting economic data in primary care. Family Practice. 2005 Jun;22(3):323-7. PubMed PMID: ISI:000229113600017. English.
6. Byford S, Leese M, Knapp M, Seivewright H, Cameron S, Jones V, et al. Comparison of alternative methods of collection of service use data for the economic evaluation of health care interventions. Health Econ. 2007 May;16(5):531-6. PubMed PMID: 17001749.
7. Petrou S, Murray L, Cooper P, Davidson LL. The accuracy of self-reported healthcare resource utilization in health economic studies. Int J Technol Assess Health Care. 2002 Summer;18(3):705-10. PubMed PMID: 12391960. Epub 2002/10/24. eng.
8. Chapko MK, Liu CF, Perkins M, Li YF, Fortney JC, Maciejewski ML. Equivalence of two healthcare costing methods: bottom-up and top-down. Health Econ. 2009 Oct;18(10):1188-201. PubMed PMID: 19097041. Epub 2008/12/20. eng.
9. Tan SS, Rutten FF, van Ineveld BM, Redekop WK, Hakkaart-van Roijen L. Comparing methodologies for the cost estimation of hospital services. Eur J Health Econ. 2009 Feb;10(1):39-45. PubMed PMID: 18340472. Epub 2008/03/15. eng.
10. Geue C, Lewsey J, Lorgelly P, Govan L, Hart C, Briggs A. Spoilt for choice: implications of using alternative methods of costing hospital episode statistics. Health Econ. 2012 Oct;21(10):1201-16. PubMed PMID: 21905152. Epub 2011/09/10. eng.
11. Chapko MK, Hedrick S. Cost as a study outcome: sensitivity of study conclusions to the method of estimating cost. Med Care. 1999 Apr;37(4 Suppl Va):AS37-44. PubMed PMID: 10217383. Epub 1999/04/27. eng.
12. Knapp M, Beecham J. Reduced list costings: examination of an informed short cut in mental health research. Health Econ. 1993;2(4):313-22.
13. Whynes DK, Walker AR. On approximations in treatment costing. Health Econ. 1995;4(1):31-9.
14. Shavit O, Leshno M, Goldberger A, Shmueli A, Hoffman A. It's Time to Choose the Study Design!: Net Benefit Analysis of Alternative Study Designs to Acquire Information for Evaluation of Health Technologies. Pharmacoeconomics. 2007;25(11):903-11. PubMed PMID: 17960950. Epub 2007/10/27. eng.
15. Pratt J, Raiffa H, Schlaifer R. Introduction to Statistical Decision Theory. Cambridge, MA: Massachusetts Institute of Technology; 1995.
16. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999 Jun;18(3):341-64. PubMed PMID: 10537899. Epub 1999/10/28. eng.
17. Claxton K. Bayesian approaches to the value of information: implications for the regulation of new pharmaceuticals. Health Econ. 1999;8(3):269-74.
18. Wilson EC. A practical guide to value of information analysis. Pharmacoeconomics. 2015 Feb;33(2):105-21. PubMed PMID: 25336432.

19. Raiffa H, Schlaifer R. Applied Statistical Decision Theory. Boston, MA: Harvard Business School; 1961.
20. Craig BA, Black MA. Incremental cost-effectiveness ratio and incremental net-health benefit: two sides of the same coin. Expert Rev Pharmacoecon Outcomes Res. 2001 Oct;1(1):37-46. PubMed PMID: 19807506. Epub 2001/10/01. eng.
21. Arrow KJ, Lind RC. Uncertainty and Evaluation of Public Investment Decisions. Am Econ Rev. 1970;60(3):364-78. PubMed PMID: ISI:A1970Y291200011. English.
22. Cachon Gr, Terwiesch C. Matching supply with demand : an introduction to operations management. 3rd ed. New York, NY: McGraw-Hill; 2013. xvii, 507 p. p.
23. Oakley JE, O'Hagan A. SHELF: The Sheffield Elicitation Framework (version 2.0). Sheffield, UK: Schoo of Mathematics and Statistics, University of Sheffield, 2010.
24. O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, D.J. J, et al. Uncertain Judgements: Eliciting Experts' Probabilities. Chichester, West Sussex: Wiley; 2006.
25. Willan A, Briggs A. Power and Sample Size Determination: The Value of Information Approach. Statistical Analysis of Cost-effectiveness Data. Chichester: John Wiley & Sons Ltd; 2006. p. 108-16.
26. Nelder JA, Mead R. A Simplex Method for Function Minimization. The Computer Journal. 1965 January 1, 1965;7(4):308-13.
27. Conti S, Claxton K. Dimensions of design space: a decision-theoretic approach to optimal research design. Med Decis Making. 2009 Nov-Dec;29(6):643-60. PubMed PMID: 19605884.
28. Price D, Musgrave S, Wilson E, Sims E, Shepstone L, Blyth A, et al. A pragmatic single-blind randomised controlled trial and economic evaluation of the use of leukotriene receptor antagonists in primary care at steps 2 and 3 of the national asthma guidelines (ELEVATE study). Health Technol Assess. 2011 May;15(21):1-132. PubMed PMID: 21554855. Epub 2011/05/11. eng.
29. Wilson EC, Price D, Musgrave SD, Sims EJ, Shepstone L, Murdoch J, et al. Cost Effectiveness of Leukotriene Receptor Antagonists versus Long-Acting Beta-2 Agonists as Add-On Therapy to Inhaled Corticosteroids for Asthma: A Pragmatic Trial. Pharmacoeconomics. 2010 May 6;28(7):597-608. PubMed PMID: 20446755. Epub 2010/05/08. Eng.
30. NHS Health and Social Care Information Centre. Prescription Cost Analysis, 2005 Leeds: NHS Health and Social Care Information Centre,; 2006 [cited 2012]. Available from: http://www.ic.nhs.uk/pubs/prescostanalysis2005.
31. Carlson JJ, Garrison LP, Ramsey SD, Veenstra DL. The potential clinical and economic outcomes of pharmacogenomic approaches to EGFR-tyrosine kinase inhibitor therapy in non-small-cell lung cancer. Value Health. 2009 Jan-Feb;12(1):20-7. PubMed PMID: 18647257.
32. Strong M, Oakley JE. An efficient method for computing single-parameter partial expected value of perfect information. Med Decis Making. 2013 Aug;33(6):755-66. PubMed PMID: 23275450.
33. Strong M, Oakley JE, Brennan A. Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample: A Nonparametric Regression Approach. Med Decis Making. 2013 Nov 18. PubMed PMID: 24246566. Epub 2013/11/20. Eng.
34. Briggs A, Nixon R, Dixon S, Thompson S. Parametric modelling of cost data: some simulation evidence. Health Econ. 2005 Apr;14(4):421-8. PubMed PMID: 15685641. Epub 2005/02/03. eng.
35. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stat Comput. 2000 Oct;10(4):325-37. PubMed PMID: ISI:000089242200005. English.

36. Philips Z, Claxton K, Palmer S. The half-life of truth: what are appropriate time horizons for research decisions? Med Decis Making. 2008 May-Jun;28(3):287-99. PubMed PMID: 18448701. Epub 2008/05/02. eng.
37. Eckermann S, Willan AR. Time and expected value of sample information wait for no patient. Value Health. 2008 May-Jun;11(3):522-6. PubMed PMID: 18179665. Epub 2008/01/09. eng.
38. Eckermann S, Willan AR. Globally optimal trial design for local decision making. Health Econ. 2009 Feb;18(2):203-16. PubMed PMID: 18435429. Epub 2008/04/26. eng.
39. British National Formulary 49. 49 ed. London: British Medical Association & Royal Pharmaceutical Societyof Great Britain; 2005.
40. Rubin DB. Mutliple Imputation for Nonresponse in Surveys. Hoboken, New Jersey: Wiley; 2004.
41. Asthma UK. Where do we stand? Asthma in the UK today. 2004.
42. Thomas M, Murray-Thomas T, Fan T, Williams T, Taylor S. Prescribing patterns of asthma controller therapy for children in UK primary care: a cross-sectional observational study. BMC Pulm Med. 2010;10:29. PubMed PMID: 20470409. Pubmed Central PMCID: 2882363. Epub 2010/05/18. eng.
43. Soriano JB, Kiri VA, Maier WC, Strachan D. Increasing prevalence of asthma in UK primary care during the 1990s. Int J Tuberc Lung Dis. 2003 May;7(5):415-21. PubMed PMID: 12757040. Epub 2003/05/22. eng.
44. Office of National Statistics. Key Population and Vital Statistics (Correction Notices). 2004 Series VS No.31,PP1 No.27. London: Office of National Statistics., 2008.

Tables

Table 1: Summary statistics - means

Statistic	Description	Intervention	Control	Increment (Δ)
n	Sample size	175	184	
E_j	QALYs at two years	1.612	1.578	0.034
$C_{n,j}$	NHS cost (excl. drugs)	£190.53	£177.35	£13.18
$C_{d,j}^{A}$	Drug cost, process A	£665.58	£563.04	£102.54
C_j^A	Total cost, process A	£856.11	£740.39	£115.72
B_j^A	Net Benefit, process A	£7203.89	£7149.61	£56.41
$C_{d,j}^{B}$	Drug cost, process B	£801.38	£511.56	£289.82
C_j^B	Total cost, process B	£991.91	£688.91	£303.00
B_j^B	Net Benefit, process B	£7069.70	£7200.57	-£130.86

Figures subject to rounding, net benefit calculated at a value of £5000 per QALY, NHS cost perspective

Table 2: Summary statistics - variance and covariance

Statistic	Description	Equation	Intervention	Control	Increment (Δ)**
n	Sample size	-	175	184	
$\sqrt{v(e_j)}$	standard deviation, QALYs	Footnote *	0.371	0.386	0.536
$\sqrt{v(c_{nj})}$	standard deviation, non-drug costs	Footnote *	£395.46	£536.81	£666.75
$\sqrt{v(c_{dj}^A)}$	standard deviation, drugs, process A	Footnote *	£416.11	£443.34	£608.03
$\sqrt{v(c_{dj}^B)}$	standard deviation, drugs, process B	Footnote *	£516.65	£384.42	£643.97
$Cov(c_{dj}^A, c_{dj}^B)$	Sample covariance, drug costs processes A and B	Footnote †	£183,804.36	£137,185.50	
$\sqrt{v(c_j^A)}$	Standard deviation, total cost, drugs estimated using process A	Footnote *	£619.84	£846.73	£1,049.35
$\sqrt{v(b_j^A)}$	Standard deviation, net benefit, process A	Footnote *	£2,010.64	£2,356.20	£3,097.47
$\sqrt{v(E_j)}$	Standard error, QALYs	Footnote ‡	0.028	0.028	0.040
$\sqrt{v(C_{n,j})}$	Standard error, non-drug costs	Footnote ‡	£29.89	£39.57	£49.60
$\sqrt{v(C_{dj}^A)}$	Standard error, drug costs, process A	Footnote ‡	£31.46	£32.68	£45.36
$\sqrt{v(C_{dj}^B)}$	Standard error, drug costs, process B	Footnote ‡	£39.05	£28.34	£48.25
$Cov\left(C_{dj}^A, C_{dj}^B\right)$	Covariance between mean drug costs processes A and B	Eq. [4-9]	£1,056.35	£749.65	£1,805.99
$\sqrt{v(C_j^A)}$	Standard error, total costs, process A for drugs cost	Footnote ‡	£46.88	£62.48	£78.11
$\sqrt{v(\Delta B_j^A)}$	Standard error of incremental net benefit, process A for drug costs	Footnote §			£217.15
$\rho_{\Delta C_n, \Delta C_d^A}$	Correlation coefficient between ΔC_n and ΔC_d^A	Footnote ||			0.352
$\rho_{\Delta C, \Delta E}$	Correlation coefficient between ΔC and ΔE	Footnote ||			-0.036
$\rho_{\Delta C_d^A, \Delta C_d^B}$	Correlation coefficient between ΔC_d^A and ΔC_d^B	Footnote ¶			0.83

* $\sqrt{v(x_j)} = \sqrt{\frac{\sum_{i=1}^{n_j}(x_{i,j}-X_j)^2}{(n_j-1)}}$

† $Cov(x_j, y_j) = \frac{\sum_{i=1}^{n_j}(x_{i,j}-X_j)(y_{i,j}-Y_j)}{(n_j-1)}$

‡ square root of equation [8]

§ square root of equation [12]

| | re-arrangement of equation [11]

¶ $\rho_{\Delta C_d^A, \Delta C_d^B} = \frac{Cov\left(\Delta C_{Dj}^A, \Delta C_{Dj}^B\right)}{\sqrt{v(\Delta C_{dj}^A)}\sqrt{v(\Delta C_{dj}^B)}}$

*** Thus $\sqrt{v(\Delta e)}$ is 0.536 and $\sqrt{v(\Delta c_N)}$ is £666.75 etc.*

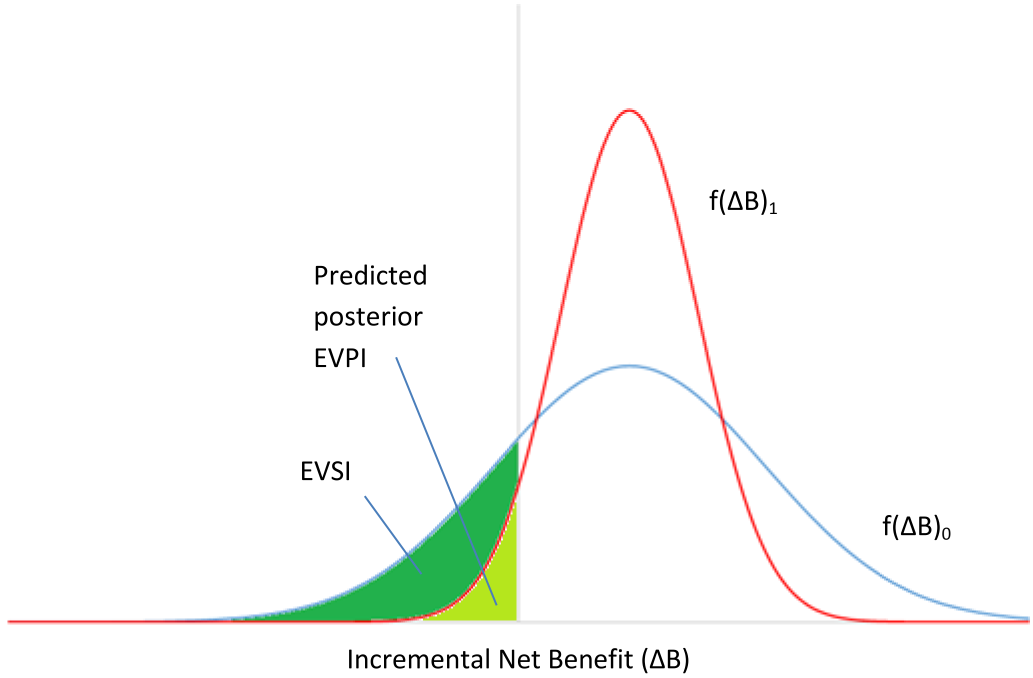
Table 3: Summary results

	EVPI	£fixed	£var	n*	EVSI	TC	OC	ENGS
INB	£32.161m	£1.305m	£288.58	2,277	£30.061m	£2.620m	£0.128m	£27.312m
QALYs	£29.228m	£1.305m	£192.39	2,473	£27.454m	£2.257m	£0.140m	£25.058m
Cost	£6.759m	£1.305m	£192.39	1,585	£ 5.738m	£1.915m	£0.089m	£ 3.733m
Non-drug Cost	£3.943m	£1.305m	£ 96.19	1,947	£ 3.070m	£1.680m	£0.110m	£ 1.280m
Drug cost (process A)	£3.433m	£1.305m	£ 96.19	1,852	£ 2.586m	£1.662m	£0.104m	£ 0.820m
Drug cost (processes A, B)	£3.433m	£1.305m	(£96.19, £9.62)	(1621, 819)	£ 2.579m	£1.633m	£0.091m	£ 0.855m
INB (QALYs, non-drug cost, drug cost process A, drug cost process B)	£32.161m	£1.305m	(£96.19, £96.19, £96.19, £9.62)	(2913, 1064, 736, 901)	£30.240m	£2.230m	£0.164m	£27.846m

EVPI =Expected Value of Perfect Information; £fixed = fixed cost of a new study; £var = variable cost of a new study; n = ENGS-maximising n per arm; EVSI = expected value of sample information; TC = total cost of new study of size n* per arm. OC = opportunity cost of patients enrolled in the 'wrong' arm of the study; ENGS = Expected Net Gain of Sampling*

Figures

Figure 1: Prior and Predicted Posterior distributions of Incremental Net Benefit



The (prior) EVPI is the sum of the shaded areas. After incorporating n predicted observations on ΔB, the prior distribution of ΔB, $f(\Delta B)_0$ is revised to $f(\Delta B)_1$. The remaining area to the left of the Y-axis is then the predicted posterior EVPI. The difference between the prior and predicted posterior EVPI is the EVSI. EVPI: Expected Value of Perfect Information; EVSI: Expected Value of Sample Information, ΔB: Incremental Net Benefit.

Figure 2: Calculation of ENGS of a trial using a combination of processes A & B

Drug costs measured using Process A

$$\Delta C_d^A = C_{d,T}^A - C_{d,C}^A$$

$$\sqrt{v(\Delta C_d^A)} = \sqrt{v(C_{d,T}^A) + v(C_{d,C}^A)}$$

Drug costs measured using Process B

$$\Delta C_d^B = C_{d,T}^B - C_{d,C}^B$$

$$\sqrt{v(\Delta C_d^B)} = \sqrt{v(C_{d,T}^B) + v(C_{d,C}^B)}$$

All other (non-drug) costs

$$\Delta C_n = C_{n,T} - C_{n,C}$$

$$\sqrt{v(\Delta C_n)} = \sqrt{v(C_{n,T}) + v(C_{n,C})}$$

Outcomes (QALYs gained)

$$\Delta E = E_T - E_C$$

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

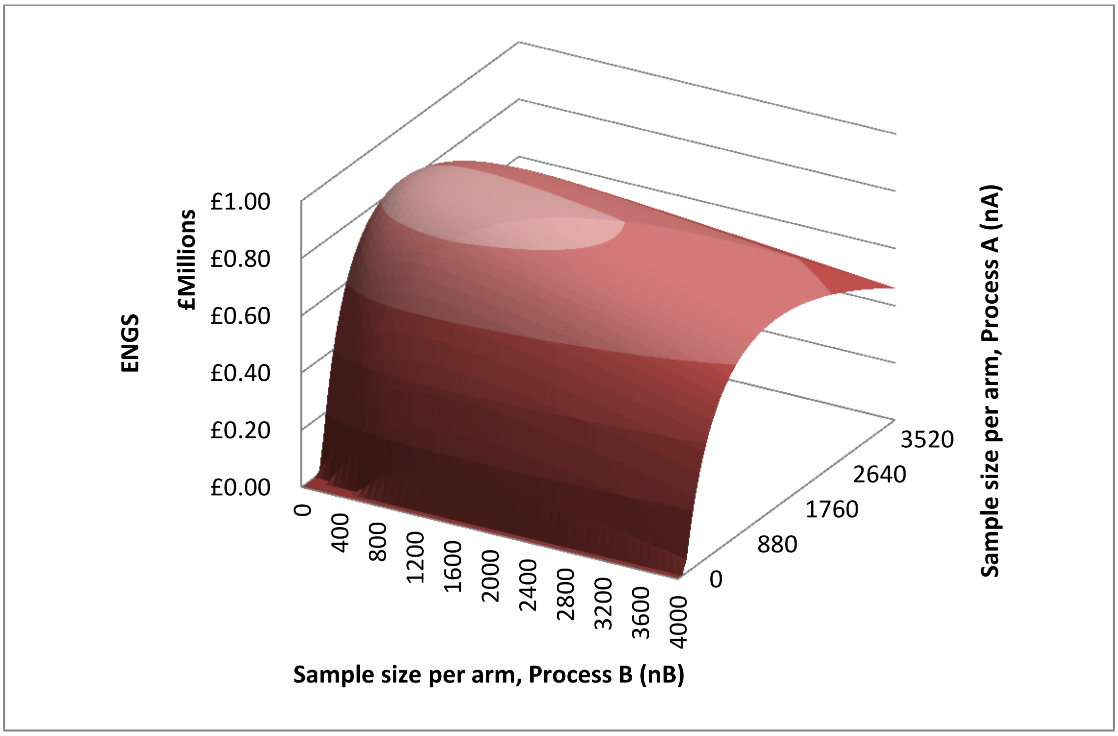
Incremental net benefit

$$\Delta B = \lambda \Delta E - \Delta C_d^A - \Delta C_n$$

$$\sqrt{v(\Delta B)} = \sqrt{\lambda^2 v(\Delta E) + v(\Delta C_d^A + \Delta C_n) - 2\lambda cov(\Delta E, \Delta C_d^A + \Delta C_n)}$$

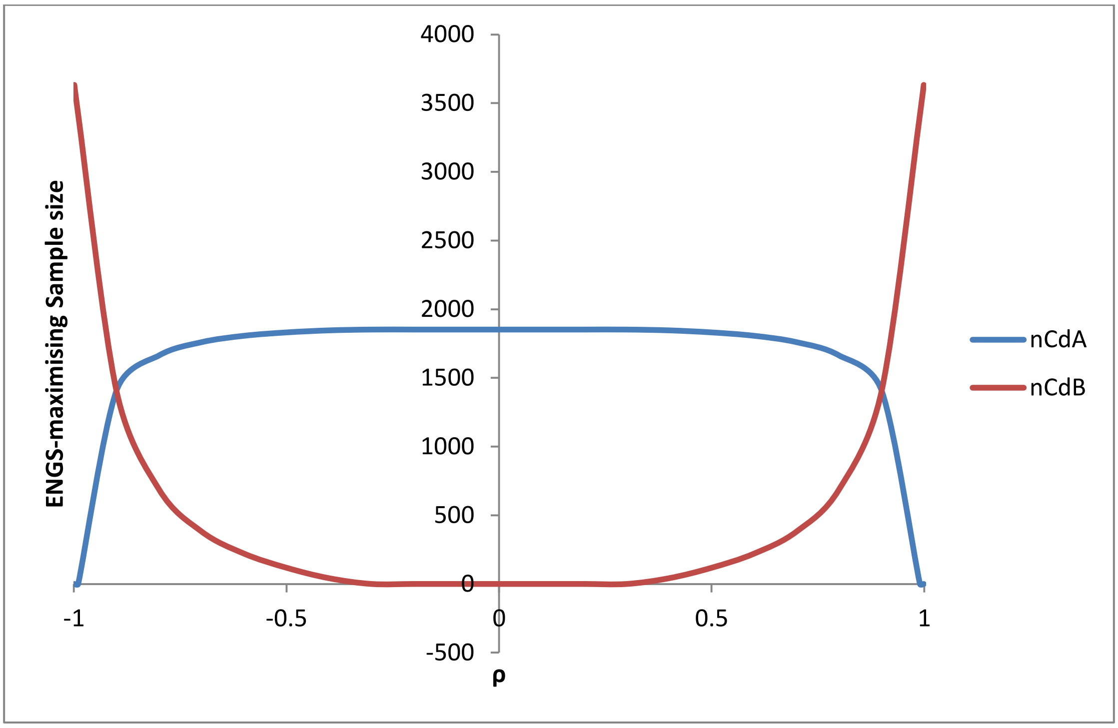
v is the variance of the mean of X and is calculated as per Equation [8]. Incremental net benefit is a function of incremental drug costs, incremental non-drug costs and incremental QALYs gained. Likewise the standard error of incremental net benefit is a function of the standard errors of incremental drug costs, incremental non-drug costs and incremental QALYs gained. If we have some information about the relationship between drug costs measured using process A and drug costs measured using process B, we can revise belief about the mean incremental cost of drugs using process A after gathering data using process B. Conceptually, this is best understood by considering extremes: If and were perfectly correlated, information on one provides perfect information about the other: the two measures are perfect substitutes and so gathering data using process B can be used to directly revise belief about plausible values obtained from process A. If there was no correlation between the two measures, gathering information on process B provides no information about process A, and therefore there is no reason to revise beliefs about plausible values for process A given data on B. Where the correlation is imperfect, we should revise beliefs in a proportionate manner, as per the algebra presented in this manuscript.

Figure 3: Optimal mix of observations from each data process



ENGS is maximised at (nA, nB) = (1621, 819)

Figure 4: Optimal mix of observations on each process as a function of ρ



Efficient Research Design: using Value of Information Analysis to estimate the optimal mix of top-down and bottom-up costing approaches in an economic evaluation alongside a clinical trial

Appendices

Edward CF Wilson, PhD[1,2]
Miranda Mugford, PhD[2]
Garry Barton, PhD[2]
Lee Shepstone, PhD[3]

1 Cambridge Centre for Health Services Research, Institute of Public Health, University of Cambridge, CB2 0SR

2 Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ, UK

3 Medical Statistics Group, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ.

ed.wilson@medschl.cam.ac.uk

Appendix 1: Details of data set, beneficial population and cost of research

The data used in the example are taken from the ELEVATE study, a study of leukotriene receptor antagonists compared with conventional treatment in asthma patients.(28) The study comprised two separate trials, on 'step 2' and 'step 3' patients. The data used here are those relating to the more severe 'step 3' patients,(29) reanalysed from the perspective of the UK NHS to divide costs into drug (C_d) and NHS-non drug items (C_n, comprising primary, secondary and tertiary resource use) at two years, and outcomes as QALYs (E) gained at two years (costs and QALYs incurred in year two were discounted at 3.5%). Incremental net benefit (ΔB) was calculated at a threshold of £5000 per QALY in order to illustrate the method demonstrated in this manuscript.

Drug cost in the original trial analysis was calculated based on individual items with the unit cost per item extracted from the British National Formulary (BNF) 2005(39) using unique BNF code at the individual preparation level. There were 27,028 items of data in the raw dataset extracted from the study database, representing individual prescription items dispensed to 683 patients over two years enrolled in the two trials comprising the ELEVATE study. The cost for each datum was recalculated at the BNF chapter section level, using aggregate cost per prescription as reported in the Prescription Cost Analysis 2005.(30) For eight observations, no sub-paragraph or chapter section data were available. Four of these were costs for specific wound dressings so the original unit cost included was applied to both summary cost estimates. The other four were blank entries that were subsequently excluded from all analyses. Therefore every patient had two estimates of drug costs over the two year study period: one bottom-up, based on actual prescribed doses of drugs (process A yielding C_d^A) and the other top-down, aggregated at the BNF section level (process B yielding C_d^B). Complete drug data were available on all patients.

As stated, the other data items were NHS non-drug cost (C_n) and QALYs gained (E) at two years. 47 (6.9%) and 283 (41.4%) of 683 observations on NHS cost and QALYs were missing. Multiple imputation was performed on the missing data including step, group, sex, age, education and employment status as coefficients. Five iterations were calculated and the results combined using Rubin's rules.(40) Data on the step 2 patients was discarded.

Population able to benefit from the information

In 2004 there were an estimated 5.2m people with asthma in the UK.(41) Approximately 7% of adolescents with asthma are at 'step 3' (requiring add-on therapy, usually of a long acting beta agonist, LABA).(42) In the absence of general population data, we assume the same proportion of adults is at step 3. During the period 1990 - 1998, GPRD data suggests the prevalence of asthma in the UK general population rose from approximately 3% to 5%.(43) This equates to an increase of approximately 0.025% per annum. Assuming a linear increase, and based on a UK population in 2004 of 59,834,300,(44) the estimated prevalence of step 3 patients in 2011 is approximately 437,297, with an incidence of 10,471 each year. Over a ten year period therefore, the potential population who could benefit from the information yielded by the studies proposed in this paper is 542,007 or 524,380 (discounted at 3.5%; Table A-1).

Table A-1: Potential Beneficial Population

Year	Patients	df	discounted
0	437297	1.000	437297
1	10471	0.966	10117
2	10471	0.934	9775
3	10471	0.902	9444
4	10471	0.871	9125
5	10471	0.842	8816
6	10471	0.814	8518
7	10471	0.786	8230
8	10471	0.759	7952
9	10471	0.734	7683
10	10471	0.709	7423
	542007		524380

Cost of research

Table A-2 summarises the predicted expenditure on a new 'ELEVATE' trial in 2010£. Overall cost is divided into a total fixed cost estimate of £1,305,470 and variable costs of £198,253 for 687 patients, or £289 per patient for a new trial.

Table A-2: Estimated budget based on expenditure for previous trial 2010£

	% FTE	months	year 1	year 2	year 3	year 4	total
Fixed costs *(including RA x1.5, admin, consumables, IT, statistical support, project supervision, expenses and overheads)*			295619	288452	292732	428854	1305470
Variable costs *(Including practice visits, practice RAs & GP costs*			69522	66096	62662		198253
Grand total							1503723

Appendix 2: Details of VoI calculations

Boxes below show the details of calculations of population EVPI and ENGS at the ENGS-maximising sample sizes. Optimal sample sizes were identified by calculating for a wide range of sample sizes using Microsoft Excel. Search algorithms were employed to identify the optimal mix of observations in Boxes A2.11 (a bespoke algorithm comparing many computations simultaneously) and A2.12 (Nelder-Mead algorithm). A Microsoft Excel spreadsheet is available on request from the corresponding author detailing all calculations.

Box A2.1: Population EVPI

Population EVPI	$P_{EVPI} = N.\sqrt{v(\Delta B)_{s,n}}.L_{N*}\left(\Delta b_0, \sqrt{v(\Delta B)_{s,n}}\right) = N.\sqrt{v(\Delta B)_{s,n}}.\left(\phi\left(\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - \frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\left[\Phi\left(-\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - I\{\Delta B_0 < 0\}\right]\right)$
Expected reduction in standard error of incremental net benefit	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1} = \sqrt{217.15^2 - 0^2} = 217.15$
Therefore population EVPI	$= 524{,}380 * 217.15 * (\phi(0.26) - 0.26[\Phi(-0.26) - 0])$ $= 524{,}380 * 217.15 * 0.28$ $= £32{,}161{,}096$

Box A2.2: ENGS, INB at n=2277 per arm

Expected net benefit of sampling n observations per arm	$ENBS(n) = (N - 2n).\sqrt{v(\Delta B)_{s,n}}.L_{N*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) - (K_s + 2k_s n + n\Delta B_0)$
Beneficial population	$(N - 2n) = (524{,}380 - 2 * 2277) = 519{,}826$
Expected reduction in standard error of incremental net benefit	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{217.15^2 - \frac{1}{\frac{1}{217.15^2} + \frac{2{,}277}{3{,}097.47^2}}} = 208.06$
Normalised mean at which to calculate unit normal loss	$\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}} = \frac{56.41}{208.06} = 0.27$
Unit normal loss	$L_{N*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) = \phi(0.27) - 0.27[\Phi(-0.27) - 0]$ $= 0.28$
Cost of sampling	$K_s + 2k_s n + n\Delta B_0$ $= £1{,}305{,}470 + 2 * 2277 * 288.58 + 2277 * 56.41$ $= £2{,}748{,}107$
Expected net benefit of sampling	$\therefore ENBS(2277) = 519{,}826 * 208.06 * 0.28 - 2{,}748{,}107$ $= £27{,}312{,}499$

Box A2.3: EVPPI, QALYs

Population expected value of partial perfect information on QALYs	$P_{EVPPI_{QALYs}} = N\sqrt{v(\Delta B)_{s,n}}.L_{N^*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right)$ $= N.\sqrt{v(\Delta B)_{s,n}}.\left(\phi\left(\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - \frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\left[\Phi\left(-\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - I\{\Delta B_0 < 0\}\right]\right)$
Pre-posterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta C)_1 = v(\Delta C)_0 = 78.11^2$ $v(\Delta E)_1 = 0$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta B)_1 = 0 + 78.11^2 - 0 = £6101.77$
Expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1} = \sqrt{£47{,}155.03 - £6101.77}$ $= \sqrt{£41{,}053.26} = £202.62$
EVPPI	$P_{EVPPI_{QALYs}} = 524{,}380 * 202.62 * \left(\phi\left(\frac{56.41}{202.62}\right) - \frac{56.41}{202.62}\left[\Phi\left(-\frac{56.41}{202.62}\right) - I\{56.41 < 0\}\right]\right)$ $= 524{,}380 * 202.62 * 0.28$ $= £29{,}228{,}197$

Box A2.4: ENGS, QALYs @ n=2473 per arm

Expected net gain of sampling data on QALYs, n=2473 per arm	$ENGS(n) = (N - 2n)\sqrt{v(\Delta B)_{s,n}}.L_{N*}\left(\Delta b_0, \sqrt{v(\Delta B)_{s,n}}\right) - (K_s + 2k_s n + n\Delta B_0)$
Beneficial population	$(N - 2n) = (524{,}380 - 2 * 2{,}473) = 519{,}434$
Expected reduction in standard error of incremental net benefit	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1}$

Preposterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta C)_1 = v(\Delta C)_0 = 78.11^2$ $v(\Delta E)_1 = \left({}^{1}/_{v(\Delta E)_0} + {}^{\mathrm{n}}/_{v(\Delta e)}\right)^{-1} = \left({}^{1}/_{0.04^2} + {}^{2473}/_{0.536^2}\right)^{-1}$ $= 0.0001082$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta b)_1 = 5000^2 * 0.0001082 + 78.11^2$ $- 2\lambda(-0.036)\sqrt{0.0001082}\sqrt{78.11^2} = £9{,}095.06$
Therefore expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{47{,}155.03 - 9095.06} = 195.09$
Normalised mean at which to calculate unit normal loss	$\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}} = \frac{56.41}{195.09} = 0.29$
Unit normal loss	$L_{N*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right)$ $= \left(\phi\left(\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - \frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\left[\Phi\left(-\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - I\{\Delta B_0 < 0\}\right]\right)$ $= 0.27$
Cost of sampling	$K_s + 2k_s n + n\Delta B_0 = £1{,}305{,}470 + 2 * 2473 * 192.39 + 2473 * 56.41$ $= £2{,}257{,}008$
Expected net benefit of sampling	$\therefore ENBS(2{,}473) = 519434 * 195.09 * 0.27 - 2{,}257{,}008$ $= £25{,}057{,}882$

Box A2.5: EVPPI, Cost

Population expected value of partial perfect information on cost	$P_{EVPPI_{Cost}} = N\sqrt{v(\Delta B)_{s,n}}.L_{N^*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right)$ $= N.\sqrt{v(\Delta B)_{s,n}}.\left(\phi\left(\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - \frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\left[\Phi\left(-\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - I\{\Delta B_0 < 0\}\right]\right)$
Pre-posterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta E)_1 = v(\Delta E)_0 = 0.04^2$ $v(\Delta C)_1 = 0$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta B)_1 = 5000^2 0.04^2 + 0 - 0 = £39{,}940.91$
Expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1} = \sqrt{£47{,}155.03 - £39{,}940.91}$ $= \sqrt{£7{,}214.12} = £84.94$
EVPPI	$= 524{,}380 * 84.94 * \left(\phi\left(\frac{56.41}{84.94}\right) - \frac{56.41}{84.94}\left[\Phi\left(-\frac{56.41}{84.94}\right) - I\{56.41 < 0\}\right]\right)$ $= 524{,}380 * 84.94 * 0.15$ $= £6{,}758{,}658$

Box A2.6: ENGS, Cost @ n=1585 per arm

Expected net gain of sampling data on cost, n=1585 per arm	$ENGS(n) = (N - 2n)\sqrt{v(\Delta B)_{s,n}}.L_{N*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) - (K_s + 2k_s n + n\Delta B_0)$
Beneficial population	$(N - 2n) = (524{,}380 - 2 * 1585) = 521{,}210$
Expected reduction in standard error of incremental net benefit	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1}$
Preposterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta E)_1 = v(\Delta E)_0 = 0.040^2$ $v(\Delta C)_1 = \left({}^{1}/_{v(\Delta C)_0} + {}^{n}/_{v(\Delta c)}\right)^{-1}$ $= \left({}^{1}/_{78.11^2} + {}^{1585}/_{1049.35^2}\right)^{-1} = £623.71$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta B)_1 = 5000^2 * 0.040^2 + 623.71 - 2\lambda(-0.036) * 0.040\sqrt{623.71}$ $= £40{,}920.26$

Therefore expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{47{,}155.03 - 40{,}920.26} = \sqrt{6234.77} = 78.96$
Normalised mean at which to calculate unit normal loss	$\frac{\Delta B_0}{\sqrt{v(\Delta B)_{s,n}}} = \frac{56.41}{78.96} = 0.71$
Unit normal loss	$L_{N*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) = (\phi(0.71) - 0.71[\Phi(-0.71) - 0]) = 0.14$
Cost of sampling	$K_s + 2k_s n + n\Delta B_0$ $= £1{,}305{,}470 + 2 * 1{,}585 * 192.39 + 1{,}585 * 56.41$ $= £2{,}004{,}746$
Expected net benefit of sampling	$\therefore ENBS_{Cost}(1{,}585) = 521{,}210 * 78.96 * 0.14 - 2{,}004{,}746$ $= £3{,}732{,}954$

Box A2.7: EVPPI, non-drug cost

Population expected value of partial perfect information on non-drug cost	$P_{EVPPI_{Non-drug\ cost}} = N\sqrt{v(\Delta B)_{s,n}}.L_{N^*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right)$ $= N.\sqrt{v(\Delta B)_{s,n}}.\left(\phi\left(\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - \frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\left[\Phi\left(-\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - I\{\Delta B_0 < 0\}\right]\right)$
Pre-posterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta E)_1 = v(\Delta E)_0 = 0.04^2$ $v(\Delta C)_1 = v(\Delta C_n)_1 + v\left(\Delta C_d^A\right)_1 + 2\rho_{\Delta C_n,\Delta C_d^A,1}\sqrt{v(\Delta C_n)_1}\sqrt{v\left(\Delta C_d^A\right)_1}$ $v\left(\Delta C_d^A\right)_1 = v\left(\Delta C_d^A\right)_0 = £45.36^2$ $v(\Delta C_n)_1 = 0$ $\rho_{\Delta C_n,\Delta C_d,1} = \rho_{\Delta C_n,\Delta C_d,0} = 0.352$ $\therefore v(\Delta C)_1 = 0 + £45.36^2 + 0 = £45.36^2$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta B)_1 = \lambda^2 0.04^2 + £45.36^2 - 2\lambda * (-0.036)\sqrt{0.04^2}\sqrt{£45.36^2}$ $= £42{,}644.50$
Expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1} = \sqrt{£47{,}155.03 - £42{,}644.50}$ $= \sqrt{4510.53} = £67.16$
EVPPI	$= 524{,}380 * 67.16 * \left(\phi\left(\frac{56.41}{67.16}\right) - \frac{56.41}{67.16}\left[\Phi\left(-\frac{56.41}{67.16}\right) - I\{56.41 < 0\}\right]\right)$ $= 524{,}380 * 67.16 * 0.112$ $= £3{,}943{,}242$

Box A2.8: ENGS, non-drug cost @ n=1947

Expected net gain of sampling data on non-drug cost, n=1947 per arm	$ENGS(n) = (N - 2n)\sqrt{v(\Delta B)_{s,n}}.L_{N*}\left(\Delta b_0, \sqrt{v(\Delta B)_{s,n}}\right) - (K_s + 2k_s n + n\Delta B_0)$
Beneficial population	$(N - 2n) = (524{,}380 - 2 * 1947) = 520{,}486$
Expected reduction in standard error of incremental net benefit	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1}$
Preposterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda \rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta E)_1 = v(\Delta E)_0 = 0.040^2$ $v(\Delta C)_1 = v(\Delta C_\mathrm{n})_1 + v\left(\Delta C_\mathrm{d}^\mathrm{A}\right)_1 + 2\rho_{\Delta C_\mathrm{n},\Delta C_\mathrm{d}^\mathrm{A},1}\sqrt{v(\Delta C_\mathrm{n})_1}\sqrt{v\left(\Delta C_\mathrm{d}^\mathrm{A}\right)_1}$ $v\left(\Delta C_\mathrm{d}^\mathrm{A}\right)_1 = v\left(\Delta C_\mathrm{d}^\mathrm{A}\right)_0 = £45.36^2$ $v(\Delta C_\mathrm{n})_1 = \left({}^{1}/_{v(\Delta C_\mathrm{n})_0} + {}^{\mathrm{n}}/_{v(\Delta c_\mathrm{n})}\right)^{-1} = \left({}^{1}/_{49.60^2} + {}^{1947}/_{666.75^2}\right)^{-1} = £208.94$ $\rho_{\Delta C_\mathrm{n},\Delta C_\mathrm{d}^\mathrm{A},1} = \rho_{\Delta C_\mathrm{n},\Delta C_\mathrm{d}^\mathrm{A},0} = 0.352$ $\therefore v(\Delta C)_1 = £208.94 + £45.36^2 + 2(0.352)\sqrt{£208.94}\sqrt{£45.36^2} = £2728.32$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta B)_1 = 5000^2 * 0.040^2 + 2728.32 - 2\lambda(-0.036) * 0.040\sqrt{2728.32} = £43{,}413.04$
Therefore expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{47{,}155.03 - 43{,}413.04} = \sqrt{3741.99} = 61.17$
Normalised mean at which to calculate unit normal loss	$\frac{\lvert \Delta B_0 \rvert}{\sqrt{v(\Delta B)_{s,n}}} = \frac{56.41}{61.17} = 0.92$
Unit normal loss	$L_{N*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) = (\phi(0.92) - 0.92[\Phi(-0.92) - 0]) = 0.10$
Cost of sampling	$K_s + 2k_s n + n\Delta B_0 = £1{,}305{,}470 + 2 * 1947 * 96.19 + 1947 * 56.41 = £1{,}789{,}880$
Expected net benefit of sampling	$\therefore ENBS_{Non-drug\ cost}(9{,}456) = 520{,}486 * 61.17 * 0.10 - 1{,}789{,}880 = £1{,}279{,}698$

Box A2.9: EVPPI, Drug Cost

Population expected value of partial perfect information on non-drug cost	$P_{EVPPI_{drugs\ cost}} = N\sqrt{v(\Delta B)_{s,n}}.L_{N^*}\left(\Delta B_0,\sqrt{v(\Delta B)_{s,n}}\right)$ $= N.\sqrt{v(\Delta B)_{s,n}}.\left(\phi\left(\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - \frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\left[\Phi\left(-\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}}\right) - I\{\Delta B_0 < 0\}\right]\right)$
Pre-posterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta E)_1 = v(\Delta E)_0 = 0.04^2$ $v(\Delta C)_1 = v(\Delta C_{\mathrm{n}})_1 + v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_1 + 2\rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}}^{\mathrm{A}},1}\sqrt{v(\Delta C_{\mathrm{n}})_1}\sqrt{v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_1}$ $v(\Delta C_{\mathrm{n}})_1 = v(\Delta C_{\mathrm{n}})_0 = £49.60^2$ $v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_1 = 0$ $\rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}},1} = \rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}},0} = 0.352$ $\therefore v(\Delta C)_1 = £49.60^2 + 0 + 0 = £49.60^2$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta B)_1 = \lambda^2 0.04^2 + £49.60^2 - 2\lambda * (-0.036)\sqrt{0.04^2}\sqrt{£49.60^2}$ $= £43{,}106.96$
Expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{£47{,}155.03 - £43{,}106.96} = \sqrt{4048.07} = £63.62$
EVPPI	$= 524{,}380 * 63.62 * \left(\phi\left(\frac{56.41}{63.62}\right) - \frac{56.41}{63.62}\left[\Phi\left(-\frac{56.41}{63.62}\right) - I\{56.41 < 0\}\right]\right)$ $= 524{,}380 * 63.62 * 0.103$ $= £3{,}433{,}463$

Box A2.10: ENGS, drug cost @ n=1852

Expected net gain of sampling data on drug cost	$ENGS(n) = (N - 2n)\sqrt{v(\Delta B)_{s,n}}.L_{N*}\left(\Delta b_0,\sqrt{v(\Delta B)_{s,n}}\right) - (K_s + 2k_s n + n\Delta B_0)$
Beneficial population	$(N - 2n) = (524{,}380 - 2 * 1852) = 520{,}676$
Expected reduction in standard error of incremental net benefit	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1}$
Preposterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta E)_1 = v(\Delta E)_0 = 0.040^2$ $v(\Delta C)_1 = v(\Delta C_{\mathrm{n}})_1 + v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_1 + 2\rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}}^{\mathrm{A}},1}\sqrt{v(\Delta C_{\mathrm{n}})_1}\sqrt{v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_1}$

	$v(\Delta C_{\mathrm{n}})_1 = v(\Delta C_{\mathrm{n}})_0 = £49.60^2$ $v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_1 = \left({}^{1}\!/_{v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_0} + {}^{\mathrm{n}}\!/_{s(\Delta C_{\mathrm{d}}^{\mathrm{A}})^2} \right)^{-1}$ $= \left({}^{1}\!/_{45.36^2} + {}^{1852}\!/_{608.03^2}\right)^{-1} = £181.97$ $\rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}}^{\mathrm{A}},1} = \rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}}^{\mathrm{A}},0} = 0.352$ $\therefore v(\Delta C)_1 = £49.60^2 + £181.97 + 2(0.352)\sqrt{£49.60^2}\sqrt{£181.97}$ $= £3{,}112.92$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta B)_1 = 5000^2 * 0.040^2 + 3112.92 - 2 * 5000(-0.036)$ $* 0.040\sqrt{3112.92} = £43{,}848.33$
Therefore expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{47{,}155.03 - 43{,}848.33} = \sqrt{3306.70} = 57.50$
Normalised mean at which to calculate unit normal loss	$\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}} = \frac{56.41}{57.50} = 0.98$
Unit normal loss	$L_{N*}\left(\Delta B_0, v(\Delta B)_{s,n}\right) = (\phi(0.98) - 0.98[\Phi(-0.98) - 0]) = 0.09$
Cost of sampling	$K_s + 2k_s n + n\Delta B_0 = £1{,}305{,}470 + 2 * 1852 * 96.19 + 1852 * 56.41$ $= £4{,}247{,}740$
Expected net benefit of sampling	$\therefore ENBS_{drug\ cost}(9{,}197) = 520{,}676 * 57.50 * 0.09 - 4{,}247{,}740$ $= £819{,}728$

Box A2.11: ENGS, drug cost, two processes @ n_a=1621, n_b=819

Prior mean incremental cost of drugs from processes A and B	$\begin{bmatrix} (\Delta C_d^A)_0 \\ (\Delta C_d^B)_0 \end{bmatrix} = \begin{bmatrix} £102.54 \\ £289.82 \end{bmatrix}$
Prior variance/covariance matrix	$\boldsymbol{V}' = \begin{bmatrix} £2{,}058 & £1{,}806 \\ £1{,}806 & £2{,}328 \end{bmatrix}$
Inverse of prior matrix	$\boldsymbol{H}' = \boldsymbol{V}^{-1} = \frac{1}{2058 * 2328 - 1806^2} \begin{bmatrix} 2328 & -1806 \\ -1806 & 2058 \end{bmatrix}$ $= \begin{bmatrix} 0.0015 & -0.0012 \\ -0.0012 & 0.0013 \end{bmatrix}$
sample precision matrix	$\boldsymbol{H} = \begin{bmatrix} {}^{n_A}\!/_{v(\Delta c_d^A)} & 0 \\ 0 & {}^{n_B}\!/_{v(\Delta c_d^B)} \end{bmatrix} = \begin{bmatrix} {}^{1621}\!/_{608.03^2} & 0 \\ 0 & {}^{819}\!/_{643.97^2} \end{bmatrix}$ $= \begin{bmatrix} 0.0044 & 0 \\ 0 & 0.0033 \end{bmatrix}$
Inverse of pre-	$\boldsymbol{H}'' = \boldsymbol{H}' + \boldsymbol{H} = \begin{bmatrix} 0.0059 & -0.0012 \\ -0.0012 & 0.0033 \end{bmatrix}$

posterior var/covar matrix	
Pre-posterior var/covar matrix	$\boldsymbol{V}'' = \boldsymbol{H}''^{-1} = \frac{1}{0.0059 * 0.0033 - 0.0012^2}\begin{bmatrix} 0.0059 & -0.0012 \\ -0.0012 & 0.0033 \end{bmatrix} = \begin{bmatrix} £182.25 & £64.81 \\ £64.81 & £324.23 \end{bmatrix}$
Therefore pre-posterior variance of incremental cost of drugs using process A	$\therefore v(\Delta C_d^A)_1 = \boldsymbol{V''}_{\boldsymbol{1,1}} = £182.25$
Expected net gain of sampling data on drug cost (n_A,n_B) observations with each process per arm	$ENGS(n_A, n_B) = \big(N - 2(n_A + n_B)\big).\sqrt{v(\Delta B)_{s,n}}.L_{N^*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) - [k_{sA}n_A + k_{sB}n_B + K_s I\{n_A > 0 \cup n_B > 0\} + (n_A + n_B)\Delta B_0]$
Beneficial population	$\big(N - 2(n_A + n_B)\big) = \big(524{,}380 - 2 * (1621 + 819)\big) = 519{,}500$
Expected reduction in standard error of incremental net benefit	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1}$
Preposterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta E)_1 = v(\Delta E)_0 = 0.040^2$ $v(\Delta C)_1 = v(\Delta C_\mathrm{n})_1 + v\left(\Delta C_\mathrm{d}^\mathrm{A}\right)_1 + 2\rho_{\Delta C_\mathrm{n},\Delta C_\mathrm{d}^\mathrm{A},1}\sqrt{v(\Delta C_\mathrm{n})_1}\sqrt{v\left(\Delta C_\mathrm{d}^\mathrm{A}\right)_1}$ $v(\Delta C_\mathrm{n})_1 = v(\Delta C_\mathrm{n})_0 = £49.60^2$ $v(\Delta C_d^A)_1 = £182.25$ $\rho_{\Delta C_\mathrm{n},\Delta C_\mathrm{d}^\mathrm{A},1} = \rho_{\Delta C_\mathrm{n},\Delta C_\mathrm{d}^\mathrm{A},0} = 0.352$ $\therefore v(\Delta C)_1 = £49.60^2 + £182.25 + 2(0.352)\sqrt{£49.60^2}\sqrt{£182.25} = £3{,}113.55$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v(\Delta B)_1 = \lambda^2 * 0.040^2 + 3113.55 - 2\lambda * (-0.036) * 0.040\sqrt{3113.55} = £43{,}849.05$
Therefore expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{47{,}155.03 - 43{,}849.05} = \sqrt{3305.98} = 57.50$
Normalised mean at which to calculate unit normal loss	$\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}} = \frac{56.41}{57.50} = 0.98$
Unit normal loss	$L_{N*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) = (\phi(0.98) - 0.98[\Phi(-0.98) - 0]) = 0.09$
Cost of sampling	$k_{sA}n_A + k_{sB}n_B + K_s I\{n_A > 0 \cup n_B > 0\} + (n_A + n_B)\Delta B_0 = 96.19 * 1621 + 9.62 * 819 + £1{,}305{,}470 * 1 + (1621 + 819) * 56.41 = £1{,}724{,}528$

Expected net benefit of sampling	$\therefore ENBS(9081{,}240) = 519500 * 57.50 * 0.09 - 1{,}724{,}528$ $= 854{,}804$

Box A2.12: ENGS of overall optimal trial design $\left(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A}, n_{\Delta C_d^B}\right) = (2913, 1064, 736, 901)$

Prior mean incremental cost of drugs from processes A and B	$\begin{bmatrix} \left(\Delta C_d^A\right)_0 \\ \left(\Delta C_d^B\right)_0 \end{bmatrix} = \begin{bmatrix} £102.54 \\ £289.82 \end{bmatrix}$
Prior variance/covariance matrix	$\boldsymbol{V}' = \begin{bmatrix} £2{,}058 & £1{,}806 \\ £1{,}806 & £2{,}328 \end{bmatrix}$
Inverse of prior matrix	$\boldsymbol{H}' = \boldsymbol{V}^{-1} = \frac{1}{2058 * 2328 - 1806^2}\begin{bmatrix} 2328 & -1806 \\ -1806 & 2058 \end{bmatrix}$ $= \begin{bmatrix} 0.0015 & -0.0012 \\ -0.0012 & 0.0013 \end{bmatrix}$
Inverse of sample var/covar matrix	$\boldsymbol{H} = \begin{bmatrix} {n_A}/{v(\Delta c_d^A)} & 0 \\ 0 & {n_B}/{v(\Delta c_d^B)} \end{bmatrix} = \begin{bmatrix} {736}/{608.03^2} & 0 \\ 0 & {901}/{643.97^2} \end{bmatrix}$ $= \begin{bmatrix} 0.0020 & 0 \\ 0 & 0.0022 \end{bmatrix}$
Inverse of pre-posterior var/covar matrix	$\boldsymbol{H}'' = \boldsymbol{H}' + \boldsymbol{H} = \begin{bmatrix} 0.0035 & -0.0012 \\ -0.0012 & 0.0035 \end{bmatrix}$
Pre-posterior var/covar matrix	$\boldsymbol{V}'' = \boldsymbol{H}''^{-1} = \frac{1}{0.0031 * 0.0022 - 0.0012^2}\begin{bmatrix} 0.0031 & -0.0012 \\ -0.0012 & 0.0022 \end{bmatrix}$ $= \begin{bmatrix} £320.84 & £107.69 \\ £107.69 & £320.40 \end{bmatrix}$
Therefore pre-posterior variance of incremental cost of drugs using process A	$\therefore v(\Delta C_d^A)_1 = £320.84$
Expected net gain of sampling data on drug cost (n_A,n_B) observations with each process per arm, n_E observations on QALYs and n_n on non-drug costs	$ENGS\left(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A}, n_{\Delta C_d^B}\right)$ $= \left[N - 2 * \max\left(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A} + n_{\Delta C_d^B}\right)\right] . \sqrt{v(\Delta B)_{s,n}} . L_{N^*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right)$ $- \left(k_{s\Delta E} n_{\Delta E} + k_{s\Delta C_n} n_{\Delta C_n} + k_{s\Delta C_d^A} n_{\Delta C_d^A} + k_{s\Delta C_d^B} n_{\Delta C_d^B} + K_s I\left\{n_{\Delta E} > 0 \cup n_{\Delta C_n} > 0 \cup n_{\Delta C_d^A} > 0 \cup n_{\Delta C_d^B} > 0\right\} + \max\left(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A} + n_{\Delta C_d^B}\right)\Delta B_0\right)$
Beneficial population	$N - 2 * \max\left(n_{\Delta E}, n_{\Delta C_n}, n_{\Delta C_d^A} + n_{\Delta C_d^B}\right)$ $= (524{,}380 - 2 * \max(2913, 1064, 736 + 901))$ $= 518{,}554$

Expected reduction in standard error of incremental net benefit	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{v(\Delta B)_0 - v(\Delta B)_1}$
Preposterior variance of INB and its components	$v(\Delta B)_1 = \lambda^2 v(\Delta E)_1 + v(\Delta C)_1 - 2\lambda\rho_{\Delta E,\Delta C,1}\sqrt{v(\Delta E)_1}\sqrt{v(\Delta C)_1}$ $v(\Delta E)_1 = \left(1/_{0.04^2} + 2{,}913/_{0.536^2}\right)^{-1} = 0.000093$ $v(\Delta C)_1 = v(\Delta C_{\mathrm{n}})_1 + v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_1 + 2\rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}}^{\mathrm{A}},1}\sqrt{v(\Delta C_{\mathrm{n}})_1}\sqrt{v\left(\Delta C_{\mathrm{d}}^{\mathrm{A}}\right)_1}$ $v(\Delta C_{\mathrm{n}})_1 = \left(1/_{49.60^2} + 1064/_{666.75^2}\right)^{-1} = 357.15$ $v(\Delta C_d^A)_1 = \ddot{\boldsymbol{V}}_{\mathbf{11}} = £320.84$ $\rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}}^{\mathrm{A}},1} = \rho_{\Delta C_{\mathrm{n}},\Delta C_{\mathrm{d}}^{\mathrm{A}},0} = 0.352$ $\therefore v(\Delta C)_1 = 357.15 + 320.84 + 2(0.352)\sqrt{357.15}\sqrt{320.84}$ $= £916.38$ $\rho_{\Delta E,\Delta C,1} = \rho_{\Delta E,\Delta C,0} = -0.036$ $\therefore v_1 = \lambda^2 0.000093 + 916.38 - 2\lambda(-0.036)\sqrt{0.000093}\sqrt{916.38}$ $= £3{,}339.43$
Therefore expected reduction in standard error of INB	$\sqrt{v(\Delta B)_{s,n}} = \sqrt{£47{,}155.03 - £3{,}339.43} = \sqrt{43{,}815.61} = 209.32$
Normalised mean at which to calculate unit normal loss	$\frac{\lvert\Delta B_0\rvert}{\sqrt{v(\Delta B)_{s,n}}} = \frac{56.41}{209.32} = 0.27$
Unit normal loss	$L_{N*}\left(\Delta B_0, \sqrt{v(\Delta B)_{s,n}}\right) = (\phi(0.27) - 0.27[\Phi(-0.27) - 0]) = 0.279$
Cost of sampling	$k_{sE}n_E + k_{sC_n}n_{C_n} + k_{sC_d^A}n_{C_d^A} + k_{sC_d^B}n_{C_d^B}$ $+ K_s I\left\{n_E > 0 \cup n_{C_n} > 0 \cup n_{C_d^A} > 0 \cup n_{C_d^B} > 0\right\}$ $+ max\left(n_E, n_{C_n}, n_{C_d^A}, n_{C_d^B}\right)\Delta B_0$ $= 2 * (96.19 * 2913 + 96.19 * 1064 + 96.19 * 736$ $+ 9.62 * 901) + £1{,}305{,}470 + (2913) * 56.41$ $= £2{,}393{,}847$
Expected net benefit of sampling	$\therefore ENBS(2913{,}1064{,}736{,}901)$ $= 518{,}554 * 209.32 * 0.279 - 2{,}393{,}847$ $= £27{,}845{,}773$

Appendix 3: Additional Figures

Figure A3.1a: Prior distribution of incremental net benefit.

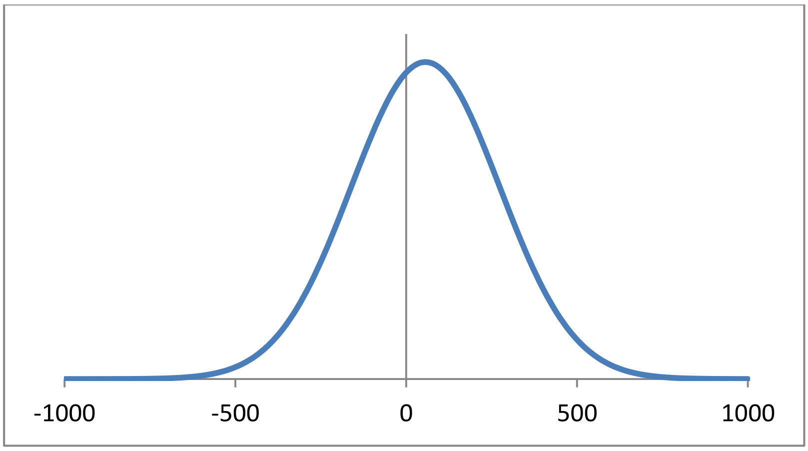


Figure A3.1b: EVSI, total cost and opportunity loss, and ENGS

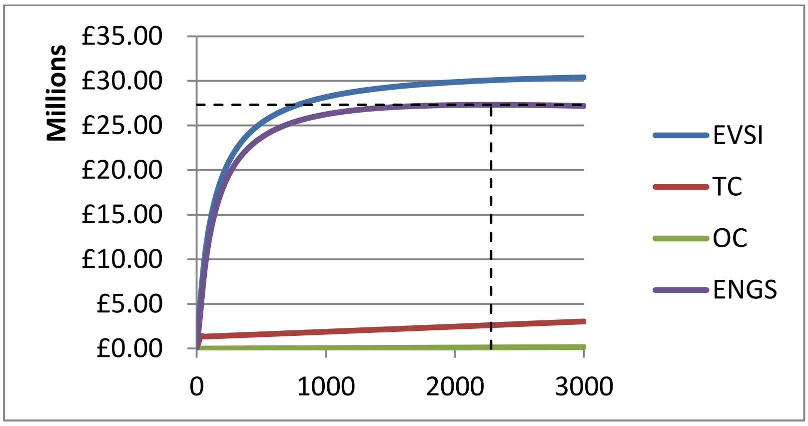


Figure A3.2a: ENGS QALYs

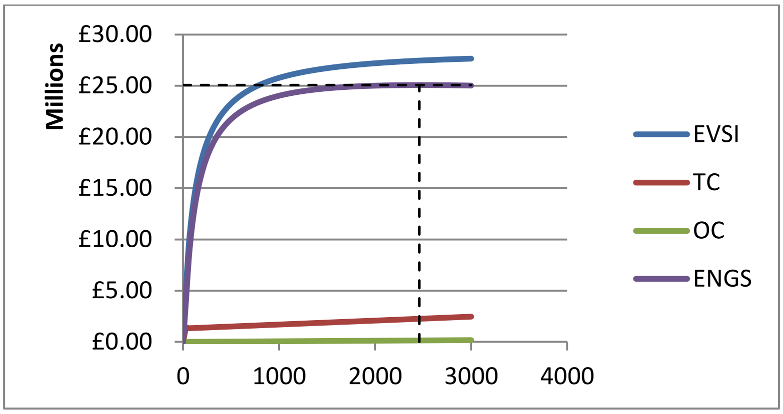


Figure A3.2b: ENGS for Cost

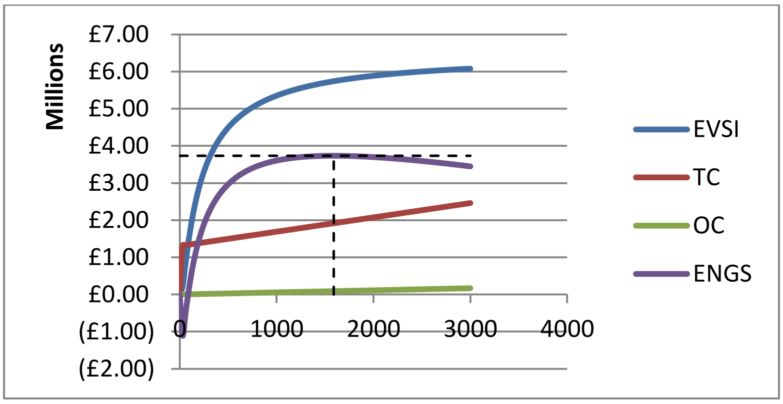


Figure A3.2c: ENGS non-drug cost

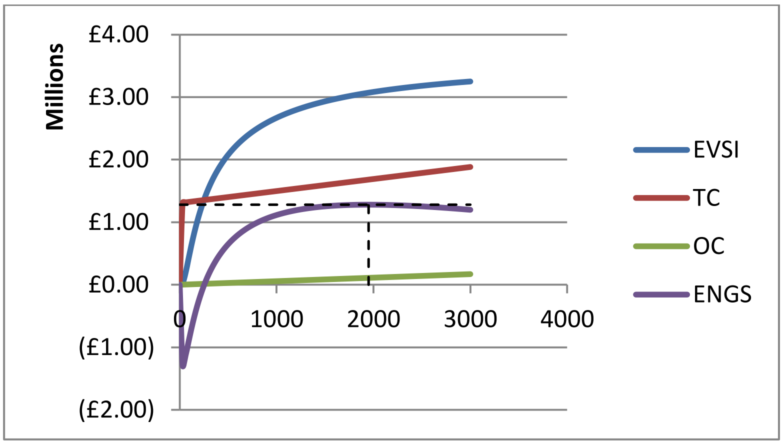


Figure A3.2d: ENGS, drug cost

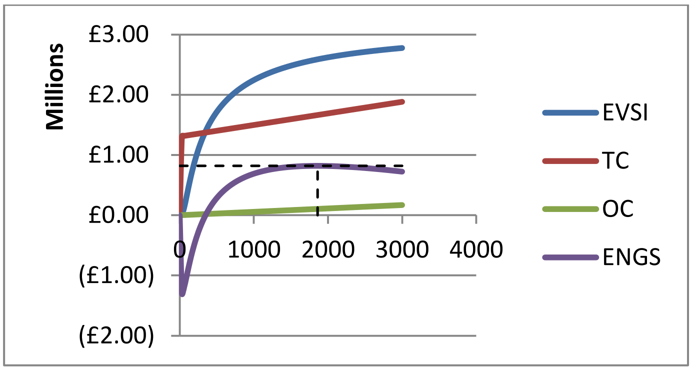


Figure A3.3: EVPI and EVPPI @ λ=£5,000