

THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS

**THE IRRELEVANCE OF INFERENCE:
A DECISION-MAKING APPROACH TO
THE STOCHASTIC EVALUATION OF
HEALTH CARE TECHNOLOGIES**

Karl Claxton

CHE Technical Paper Series 8

The Irrelevance of Inference: a Decision-Making Approach to the Stochastic Evaluation of Health Care Technologies.

Karl Claxton

Department of Economics and Related Studies,
University of York

23rd June 1997

Abstract

The literature which considers the statistical properties of cost-effectiveness analysis has focused on estimating the sampling distribution of either an incremental cost-effectiveness ratio or incremental net benefit for classical inference. However it is argued here that rules of inference are arbitrary and entirely irrelevant to the decisions which clinical and economic evaluations claim to inform. Decisions should be based only on the mean net benefits irrespective of whether differences are statistically significant or fall outside a Bayesian range of equivalence. Failure to make decisions in this way by accepting the arbitrary rules of inference will impose opportunity costs which can be measured in terms of resources or health benefits forgone. The distribution of the incremental net benefit is only relevant to deciding whether it is worth collecting more information to reduce the expected costs of basing treatment choice on existing estimates of the mean net benefit. A framework for decision making and establishing the value of additional information is presented which is consistent with the decision rules in CEA. This framework can distinguish the simultaneous but conceptually separate steps of deciding which alternatives should be chosen, given existing information, from the question of whether more information should be acquired. It mirrors the sequential nature of decision making: making an initial decision; deciding to gather evidence; revising decisions in the light of this new information; and again considering whether more information is required. It also ensures that the type of information acquired is driven by the objectives of the health care system, is valued in a way which is consistent with the budget constraint on service provision and that research is designed efficiently.

Acknowledgements: Andrew Jones, Larry Lacey, Andrew Street and Steve Walker for comments on previous drafts.

Correspondence: Karl Claxton, Lecturer in Economics, Department of Economics and Related Studies, University of York, Heslington, York, YO1 5DD. Tel: 01904 433798 e-mail kpc1@york.ac.uk.

Keywords: Cost-effectiveness analysis; Statistics; Clinical Trials;

JEL classification: I18; D61

1 Introduction

The literature which considers the statistical properties of cost-effectiveness analysis has focused on estimating the sampling distribution of incremental cost-effectiveness ratios ¹⁻⁶. This approach leads to problems of infinite value of the cost-effectiveness ratio when incremental benefits can be zero, and infinite value of the effectiveness-cost ratio when incremental cost can be zero. However an appropriate test statistic should be relevant for decision making, not simply inference, and should combine both the summary measure of cost-effectiveness and the decision rule used to establish which of the alternatives is cost-effective. The incremental cost-effectiveness ratio is an appropriate summary measure of cost-effectiveness when comparing mutually exclusive alternatives, indeed there is no legitimate role of what have been called average cost-effectiveness ratios ₁. However in the absence of dominance an incremental cost-effectiveness ratio is not sufficient to establish which alternative is cost-effective and should be implemented ⁷⁻¹⁰. An incremental cost-effectiveness ratio must be compared to the opportunity cost of implementing the more costly but more effective alternative. The opportunity cost is the health benefit foregone by incurring the incremental costs of the more effective treatment, displacing the marginal project and foregoing the benefits that it provided. This has been described as the critical cost-effectiveness ratio ¹¹⁻¹³ and is the shadow price of the budget constraint on service provision ^{14,15}, or the monetary value of health outcome implicit in existing levels of service provision ¹⁶. This decision rule can also be expressed in terms of an effectiveness-cost ratio and the new treatment ($j=2$) is cost-effective if the effectiveness-cost ratio is greater than the effectiveness-cost ratio of the marginal project which will be displaced (g): the shadow price of health outcome.

$$1a \quad (U_2 - U_1) / (C_2 - C_1) > g$$

$$1b \quad (C_2 - C_1) / (U_2 - U_1) < 1/g$$

Where U_j is the health outcome with treatment j and C_j is the cost of treatment j . This is equivalent to asking if the cost-effectiveness ratio is less than the critical ratio or the marginal willingness to pay for health outcome implicit in the existing budget constraint in 1b. No decisions can be made until the value of g has been established ₂ ^{12, 17, 18}. Following Phelps and Mushlin (1991) the decision rule in 1a and 1b can be rearranged so that the cost-effectiveness of the alternatives are expressed in incremental net benefits which can be measured on either an effectiveness (η) or a monetary scale (μ). The new treatment is cost-effective if:

$$2a \quad U_2 - g \cdot C_2 > U_1 - g \cdot C_1 \quad \text{or} \quad \eta = (U_2 - U_1) - g \cdot (C_2 - C_1) > 0$$

$$2b \quad 1/g \cdot U_2 - C_2 > 1/g \cdot U_1 - C_1 \quad \text{or} \quad \mu = 1/g \cdot (U_2 - U_1) - (C_2 - C_1) > 0$$

Clearly each of these expressions are equivalent and as described by Phelps and Mushlin suggests the "near" equivalence of CEA and CBA when the value of $1/g$ is equal to the marginal willingness to pay by particular patient groups ^{12, 19}.

An appropriate test statistic for CEA

Whenever a decision is made, or CEA is used to make a statement about what is cost-effective, a monetary valuation of health outcome is applied. In most cases this is done implicitly (it is also implicit in the selection of "relevant alternatives" during the design of any evaluative study) ^{17, 19}. However if it is made explicit then it can be applied consistently and incremental net benefit can be used as the test statistic. This can directly address the hypothesis posed by economic evaluation: that one alternative is more cost-effective than another. The incremental net benefit in 2a and 2b have been used in a Bayesian framework to establish prior and sampling distributions for CEA ²⁰, and in a classical frequentist framework to establish a test statistic for CEA ²¹.

$$\begin{aligned} 3a \quad \text{Var}(\eta) &= \text{Var}((U_2 - U_1) - g(C_2 - C_1)) \\ &= \sigma_e^2 + g^2 \cdot \sigma_c^2 - 2 \cdot g \cdot \sigma_{ce} \\ \text{where: } \sigma_e^2 &= \text{Var}(U_2 - U_1) = \text{Var}(U_2) + \text{Var}(U_1) - 2 \cdot \text{Cov}(U_2, U_1) \\ \sigma_c^2 &= \text{Var}(C_2 - C_1) = \text{Var}(C_2) + \text{Var}(C_1) - 2 \cdot \text{Cov}(C_2, C_1) \\ \sigma_{ce} &= \text{Cov}((C_2 - C_1), (U_2 - U_1)) \\ 3b \quad \text{Var}(\mu) &= \text{Var}(1/g(U_2 - U_1) - (C_2 - C_1)) \\ &= 1/g^2 \cdot \sigma_e^2 + \sigma_c^2 - 2/g \cdot \sigma_{ce} \end{aligned}$$

It can also be established that the test statistic will be equivalent when incremental net benefits are expressed in either effectiveness (4a) or monetary (4b) terms.

$$\begin{aligned}
4a \quad & \text{Ho: } \eta = 0, \quad Z_{\eta} = \frac{(\overline{U_2 - U_1}) - g \cdot (\overline{C_2 - C_1})}{\sqrt{\text{Var}(\eta)/n}} \\
4b \quad & \text{Ho: } \mu = 0, \quad Z_{\mu} = \frac{1/g \cdot (\overline{U_2 - U_1}) - (\overline{C_2 - C_1})}{\sqrt{\text{Var}(\mu)/n}} \\
Z_{\mu} &= \frac{1/g \cdot (\overline{U_2 - U_1}) - (\overline{C_2 - C_1})}{\sqrt{\text{Var}(\mu)/n}} = \frac{(\overline{U_2 - U_1}) - g \cdot (\overline{C_2 - C_1})}{\sqrt{g^2 \text{Var}(\mu)/n}} \\
&= \sqrt{\frac{g^2 (1/g^2 \cdot \sigma_u^2 + \sigma_c^2 - 2/g \cdot \sigma_{ce})}{n}} = \sqrt{\frac{(\sigma_u^2 + g^2 \sigma_c^2 - 2 \cdot g \cdot \sigma_{ce})}{n}} \\
Z_{\eta} &= Z_{\mu}, \quad \forall g
\end{aligned}$$

Clearly the mean and the variance of the incremental net benefit will be a function of the shadow price of the budget constraint.

$$5a \quad \partial \text{Var}(\eta) / \partial g = 2 \cdot g \cdot \sigma_c^2 - 2 \cdot \sigma_{ce}$$

As the budget is relaxed the cost-effectiveness ratio of the marginal project will increase (g will fall) as more effective but more costly programmes are adopted. Consequently g^2 in 3a will fall and the variance of the incremental net benefit will also fall (if $g > \sigma_{ce}/\sigma_c^2$). This is illustrated using a simple numerical example detailed in table 1 (where $\sigma_{ce}=0$), appendix A. Figure 1a illustrates the incremental net benefit of treatment t_2 valued in terms of health outcome. If the value of health outcome is £400 per outcome gained the null is rejected and treatment t_2 can be regarded significantly more cost-effective than treatment 1. Figure 1a demonstrates that sample variance is dependent on budgetary restrictions on service provision. As the budget is relaxed the confidence intervals narrow around the increasing mean incremental net benefit. The sample variance approaches σ_u^2/n as g becomes very small ($1/g \rightarrow \infty$). No weight is placed on cost and when $g=0$ the decision rule becomes a purely clinical effectiveness rule. The reverse is also true and when a cost minimisation rule is used ($1/g=0$, or $g \rightarrow \infty$), where health benefits are disregarded, sample variance is infinite. This relationship between sample variance and the shadow price of the budget constraint is reversed (if $1/g > \sigma_{ce}/\sigma_u^2$) when net benefits are measured on a monetary scale using 3b.

$$5b \quad \partial \text{Var}(\mu) / \partial 1/g = 1/g \cdot \sigma_u^2 - 2 \cdot \sigma_{ce}$$

This is illustrated in figure 1b for the same numerical example in table 1. Now the sample variance increases with the value placed on health outcome and as $1/g$ becomes very large and an infinite value is placed on health outcome, the sample variance approaches infinity and the confidence limits are unbounded (as is the incremental net benefit). Similarly when the cost minimisation rule is used ($1/g=0$) the sample variance will be σ_c^2/n . However despite the fact that the sample variance will be related to the shadow price of the budget constraint in different ways depending on whether the incremental net benefits are measured in effectiveness or monetary terms Z_{η} will always be equal to Z_{μ} and the p values of any differences in net benefits will be the same irrespective of the scale used for net benefits (because g affects both the denominator and the numerator of Z_{η} and Z_{μ}).

This can lead to an interesting situation where two non mutually exclusive alternative programmes have identical mean net benefits with identical variances for incremental cost and benefits but where different values of g result in different test statistics. This is illustrated in figure 2 using a simple numerical example detailed in table 2, appendix A. These two examples could represent the same clinical decision problem each with the same two mutually exclusive treatment alternatives considered under two different budgetary regimes ($1/g=\text{£}400$ in figure 1b and $\text{£}700$ in figure 2) with different acquisition costs for the new treatment (t_2). The new treatment in figure 1 and 2 has identical incremental net benefit when compared to an identical current practice t_1 but in figure 2 the null hypothesis

is accepted and we conclude that there is no statistically significant difference in cost-effectiveness between current practice and the new treatment, whereas the null was rejected in figure 1b. The difference lies only in terms of the context such as the budgetary regime and the value of g . These numerical examples support the argument that the explicit monetary valuation of health outcome is an essential and unavoidable element in estimating the mean net benefit and making inference about cost-effectiveness.

The shadow price of the budget constraint determines which of the alternatives will be regarded as cost-effective and establishing the value of this key parameter should not be abdicated to the vagaries of uninformed and implicit judgment of social decision makers. Establishing the value of g may be problematic and will be specific to time, place, perspective, and the measure of health outcome deemed to be appropriate^{9, 12, 14-16, 19}. However these problems cannot be avoided because even if a reliable test statistic for an incremental cost-effectiveness ratio was available it would not be possible to test a coherent hypothesis without a value for g . This is not to argue for a return to a Paretian approach or to suggest that a Paretian approach has any normative superiority over an extra-welfarist or social decision making approach²²⁻²⁵. CEA is founded on an extra-welfarist view but within this the shadow price of the budget constraint is an empirical question¹³⁻¹⁴ which is an essential element in establishing which health care programmes should be implemented. For those of us who reject a Paretian view of the world a tendency to avoid addressing directly the explicit monetary valuation of health outcome in CEA is understandable but it is also inconsistent and incoherent. The explicit monetary valuation of health outcome is an essential, unavoidable and (given the normative judgement which determine the existing budget) a positive empirical question in testing the hypothesis that a new treatment is cost-effective when compared to current practice. It seems appropriate that the decision rules and the test statistic for the hypothesis posed by cost-effectiveness analysis should reflect this.

2 The irrelevance of inference

However whether this test statistic is at all relevant to the decision of which programme should be adopted is not clear. It has been argued for some time that the level of significance or the probability of a type I error in a pragmatic clinical trial (which wishes to inform clinical decision making) is entirely irrelevant²⁶:

“If we simply wish to decide which of the two [treatments] to use it is immediately clear that the error probability α is quite irrelevant; if $A=B$ it cannot matter which we decide to choose. ...if A and B are equivalent, there is no drawback in choosing one or the other of them always provided that the assessment of the results is sufficiently broad based. ...If we are not interested in α , we no longer need to minimise it; ...we have no need to make a significance test, ...we simply choose the treatment with the better mean value;” (Schwartz D. and Lellouch J. 1967)

Indeed it is argued here that the rules of classical statistical inference are arbitrary, inconsistent with the objectives of any coherent health care system, and impose unnecessary costs. These rules should be rejected: including significance testing, confidence intervals and p values. Similarly the Bayesian counterparts to confidence intervals, in the form of ranges of equivalence²⁷, are also only concerned with inference rather than decision making, are equally arbitrary, and should also be rejected in favour of a decision making approach²⁸.

“It must be recognised that clinical trials are not there for inference but to reach a decision, and the omission of their *raison d’être* is serious. In the long term utility is realistic and, indeed necessary. ... It is only by using expected utility that we can be sure that our actions fit together sensibly. I suspect that the procedure of continuing with a trial until a tail area probability in the posterior is small is just as incoherent as a belief based on the tail area p -value. Or if it is coherent, it implies an inept utility, such as one taking only values of 0 and 1.” (Lindley 1994 page 393)

Inspired by Lindley a framework for deciding which of the competing mutually exclusive alternatives should be chosen and deciding whether it is efficient to gather more information about the decision problem to inform this choice is presented.

The feasibility and necessity of a decision making approach

The possibility of abandoning inference (based on either a frequentist or Bayesian view of probability) and taking a decision making approach has been discussed for some time but has been rejected on a number of grounds ²⁷:

“Although there have been many attempts to place clinical trials within such a decision theoretic framework, in our formulation we specifically do not include utility assessments. Our reason is that when a decision is whether or not to discontinue the trial, coupled with whether or not to recommend one treatment in preference to the other, the consequences of any particular course of action are so uncertain that they make the meaningful specification of utilities rather speculative. ...” (Spiegelhalter, Freedman, and Parmar 1994 page 360)

This leads to a somewhat contradictory position later in this influential paper by Spiegelhalter, Freedman and Parmar.

“The choice [of acceptable error] should in principle anyway, be made from decision theoretic considerations of expected utility. But we have already said that this is unrealistic.” (Spiegelhalter, Freedman, and Parmar 1994 page 370)

However even at the time a small number including Lindley and Berry argued that abandoning inference is both necessary and feasible but were not able to outline a framework which would make it possible ²⁹.

“I disagree that a decision theoretic approach using utilities is unrealistic in clinical trials: “the consequences of any particular course of action are so uncertain that they make the meaningful specification of utilities rather speculative”. Speculation and assessing uncertainty are the stuff of the Bayesian approach. In deciding whether to stop a trial the authors spurn utilities and are left with conventional benchmarks such as 2.5% and 5%. In my view, deciding when to stop a trial requires considering why we are running it in the first place, and this means assessing utilities.” (Berry 1994 page 399)

The reasons why a decision making approach was rejected in the past fall into three categories. Firstly it was not clear how health outcomes and resource costs could be measures on the same scale and therefore it was not possible to construct “utilities” which were sufficiently “broad based”. The issue of cost was subsumed into either the clinically significant difference that is worth detecting in the frequentist approach ^{20, 30, 31};

“There will be a minimum worthwhile improvement for the experimental treatment to attain before the upheaval [cost] of a change of policy is justified” (Whitehead 1992 p22),

or the range of equivalence in a Bayesian framework ²⁷;

“...the range of equivalence, which combines evidence regarding the relative differences between the treatments in toxicity, cost and inconvenience” (Spiegelhalter 1994 p368).

The discussion in section 1 and 2 above and recent work in this area ^{14, 19} demonstrates that this is no longer a problem of principle: if the budget for service provision is fixed then the shadow price of the budget constraint is a fixed parameter whose value can be established by empirical analysis. Costs can be rescaled into health outcome or health outcome into monetary terms. The decision as to “recommend one treatment in preference to the other” or “whether or not to discontinue the trial” can then simply be based on (posterior) mean incremental net benefit without reference to arbitrary rules.

Secondly, early attempts to found clinical trials in a decision theoretic framework lead to very large or unbounded predicted sample sizes ³²⁻³⁷. Since there was no way to incorporate resource cost into this analysis very large predicted sample sizes are not surprising when the marginal resource cost of an additional sample is excluded (implicitly g was zero and an infinite value was placed on the benefits of sample information). Now that resource costs can be incorporated using the shadow price of the budget constraint the marginal benefits and marginal costs of sample information can be established and unbounded sample size is not a problem ^{17, 20} (although very large sample size may still be efficient in certain circumstances).

Finally if a decision making approach is used to decide, “whether or not to discontinue the trial, coupled with whether or not to recommend one treatment in preference to the other”, there is a justifiable concern that the results of the research will not have an impact on the decisions of clinical practitioners who may continue to demand substantial clinical improvements at 5% significance and 80% power before they will adopt a new treatment. However to select some methods (reject others) and design research based on the concerns of current practitioners is rather circular and makes discussion of appropriate methods seem rather irrelevant. It implies that either the current decision making processes of practitioners in assessing and acting on evidence are optimal with regard to the objectives of the health care system or are not amenable to change. A different view is taken here: it is not assumed that current decision making is necessarily optimal and believe (based on evidence which has accumulated over the last 30 years in the economics literature and else were) that it may be amenable to change. In selecting appropriate methods the first step is to select an approach which “best” meets the objectives of the health care system (a normative framework) then conduct positive research to establish what incentives or institutional arrangements are required to persuade decision makers to act on this evidence in the way intended. Here the former is addressed and it is left to others to continue to address the latter.

3 The choice between mutually exclusive alternatives

If the objective is to maximise health gain for a given budget then programmes should be selected based on the (posterior) mean net benefit irrespective of whether any differences are regarded as statistically significant or fall outside a Bayesian range of equivalence. This is because one of the mutually exclusive alternatives must be chosen and this decision can not be deferred. In this example by selecting current practice t_1 we reject t_2 and visa versa. The opportunity cost of failing to make the correct decision based on the mean are symmetrical and the historical accident that dictates which of the alternatives is regarded as current practice is irrelevant.

This can be demonstrated using the numerical example which was illustrated figure 2 and detailed table 2. When the value of health outcome is £700 per unit of health outcome the incremental net benefit of t_2 is valued at £400 or 0.5714 units of health outcome for each patient treated. This incremental net benefit is not statistically significant with a p-value of 0.1218 and according to the rules of inference the null is accepted and the new treatment is rejected (current practice t_1 is chosen). However this arbitrary rule imposes unnecessary costs on individual patients and on the population of current and future patients who enter this decision problem (the same argument also applies to clinical evidence about efficacy and effectiveness which are simply a special case where $g=0$). For example; assume an incidence of 2,000 patients entering this decision problem each year over 10 years. Applying a discount rate of 6% the opportunity costs of failing to adopt the new treatment simply because the difference in net benefit is not statistically significant can be valued at £6,241,354 or at 8,916 units of health outcome. In this example when $1/g=£300$ the incremental net benefit of t_2 is -£400 and current practice should be chosen. However if the new treatment is adopted the opportunity cost can also be valued at £6,241,354 (the opportunity costs of these two equally arbitrary decisions are the same). No one would ever suggest that a new treatment should be adopted when its mean net benefit is less than current practice but in terms of opportunity cost this is precisely what is implied by the arbitrary rules of inference.

The measures of net benefits presented in these numerical examples are based on a particular objective or social welfare function which may be judged inappropriate and consequently the measures of net benefit may be regarded as incomplete. However if there are equity issues which need to be incorporated they can be made explicit with appropriate adjustments to the measure of outcome³⁸. Similarly on the issue of the safety of a new technology if there is particular concern for rare but catastrophic events then these undesirable outcomes should be given appropriate weight in the calculation of expected net benefits⁶. If decision makers wish to adopt some type of voting rule then the decision may focus on the median rather than the mean net benefits and the treatment which benefits the greatest number could be adopted^{4,38}. Attitudes to risk can be incorporated in the measures of outcome but if we wish to incorporate the fact that some individuals preferences violate the axioms of expected utility theory then prospect theory or some notion of regret can be used in the measure of net benefits³⁹⁻⁴¹. It is not necessary to take a position on the appropriate social welfare function; different definitions of need; the normative content of expected utility theory; or the best way to incorporate important and legitimated equity concerns (although it is worth noting that they will all have implications for the measures of outcome and the design of studies not just their interpretation) only to note that these issues do not change the fundamental point that inference is irrelevant and the decisions must still be based on either the mean or median of the distribution. Whatever view is taken, confidence intervals, p values and levels of significance still remain entirely irrelevant to the decision.

4 The decision to acquire additional information

The distribution of net benefits is entirely irrelevant to the choice between mutually exclusive alternatives, which (given existing information) should be based on the mean. The distribution of net benefits is only relevant to the decision of whether to collect more information about t_2 and t_1 to inform this treatment choice now and in the future. To illustrate this decision making approach a fixed sample design is taken where all the sample information will be available at the end of any trial, and any sample information will take some time to acquire. This means that there will be a sequence of decisions: making an initial treatment decision (based on prior information); deciding to gather evidence; revising decisions in the light of this new information (based on the posterior mean); and again considering whether more information is required (based on the new prior mean). The approach can distinguish the simultaneous but conceptually separate steps of deciding which alternatives should be chosen, given existing information, from the question of whether more information should be acquired. It mirrors the sequential nature of decision making and ensures that the type of information acquired (research) is driven by the objectives of the health care system and is valued in a way which is consistent with the budget constraints on service provision. This is illustrated using the same numerical examples discussed above. For comparability we assume that the original prior was uninformative in which case the sample means in figure 1 and 2 are also posterior means and the sample variances (in tables 1 and 2) are also the posterior variances. However once a treatment choice has been made based on existing information we must consider whether to acquire more information by collecting additional sample information (another trial). The posterior (sample) mean and variance from the original trial become the prior mean and variance when facing the next choice in this sequence of decisions.

The Expected Value of Perfect Information

The treatment decision is uncertain because net benefits are stochastic. The expected cost of this uncertainty will be determined by the probability that a decision based on mean net benefit will be wrong and the opportunity loss if the wrong decision is made. The expected opportunity loss²⁶ is the expected cost of uncertainty surrounding the treatment decision when it is based on existing (prior) information. It can also be interpreted as the expected value of perfect information (EVPI), since perfect information (an infinite sample) would eliminate the possibility of making the wrong decision^{20, 42-44}.

$$6a \quad EVPI = \frac{2}{g} \cdot \sigma_0 \cdot L(D_0), \quad \text{where} \quad D_0 = \frac{|\eta_0 - \eta_b|}{\sigma_0}$$

$$L(D_0) = \text{unit normal loss integral for } D_0$$

The probability that the decision will be wrong is determined by the distance of the prior mean incremental net benefits of t_2 (η_0) from the point of indifference between t_1 and t_2 (where $\eta_0 = \eta_b = 0$) and the prior distribution of η_0 ($\text{Var}(\eta_0) = \sigma_0^2$). The opportunity cost (loss) if the decision is wrong can be established using a loss function where opportunity loss is simply the difference in net benefit between the best choice and the choice actually made (and will be minimised when decisions are based on the mean net benefits). The prior incremental net benefit of t_2 can be expressed in monetary terms as $k_2 \cdot \eta_0$ (where $k_2 = 1/g$) and the prior incremental net benefit of t_1 can be expressed in money terms as $k_1 \cdot \eta_0$ (where $k_1 = -1/g$). The slope of the loss function is therefore $|k_2 - k_1| = 2/g$ and represents the value placed on opportunity losses when they are incurred.

The EVPI is the maximum value that can be placed on additional information to inform treatment choice for an individual patient. However any information acquired will be non rival and can be used to inform the treatment decision for all eligible patients entering the same clinical decision problem now and in the future. Given an estimate of the incidence of patients entering this problem (I) in each period (t) we can establish the EVPI for a population:

$$6b \quad EVPI \sum_{t=1}^T \frac{I_t}{(1+r)^t}$$

In the examples discussed earlier we assumed 2,000 patients entering each year over ten years at a discount rate (r) of 6%. The EVPI for these two examples over a range of values of $1/g$ is illustrated in figure 3. In example 2, where

the incremental net benefit was not regarded as significant, the EVPI is much higher (£643,194 when $1/g = £700$) than in example 1 where it was (£46,570 when $1/g = £400$). The difference between example 1 and 2 is not the decision about which treatment to adopt but the value of acquiring more information to inform this decision in the future. Since the EVPI represents the maximum value of additional information (clinical research) it can be used to eliminate proposed research where the estimated costs exceed these maximum benefits²⁰. In these examples if the fixed cost of clinical research was estimated to be £50,000 then we would not wish to acquire more information in example 1 but additional sample information is potentially cost-effective in example 2. The EVPI for both examples is closely related to the shadow price of the budget constraint. In both cases the EVPI reaches a maximum when $1/g$ is equal to the prior incremental cost effectiveness ratio or where $\eta_0 = \eta_b = 0$. At this point the decision maker would be indifferent between t_1 and t_2 and the chance that the decision based on the prior mean will be wrong reaches a maximum. When $1/g$ is greater than the prior incremental CER any increase in its value will increase the value placed on opportunity losses when they occur but will also reduce the probability of incurring opportunity losses because $|\eta_0 - \eta_b|$ will increase. In these examples the later offsets the former but this need not be the case and will depend on the strength of prior information¹⁰.

The Expected Net Benefits of Sample Information (fixed sample allocation)

The EVPI provides a necessary but not sufficient condition for choosing to acquire more sample information. To decide whether acquiring additional information will be efficient it is necessary to establish the marginal benefit and the marginal cost of sampling. The benefit of additional sample information is the reduction in the cost of uncertainty surrounding the treatment decision. For an individual patient the expected benefits of a sample (n) allocated between the experimental ($n_{(2)}$) and control arm ($n - n_{(2)}$) of the trial can be expressed as the expected value of sample information ($EVSI|n, n_{(2)}$). Initially we assume a fixed and equal allocation of the sample to experimental and control ($n_{(2)} = n/2$)²⁰.

$$7a \quad EVSI|n, n_{(2)} = \frac{2}{g} \cdot \sqrt{V|n, n_{(2)}} \cdot \sigma_0 \cdot L(D|n, n_{(2)})$$

$$\text{where } D|n, n_{(2)} = \frac{|\eta_0 - \eta_b|}{\sigma_0 \cdot \sqrt{V|n, n_{(2)}}}, \quad V|n, n_{(2)} = \frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma_{(2)}^2}{n_{(2)}} + \frac{\sigma_{(1)}^2}{n - n_{(2)}}}$$

Where $\sigma_{(2)}^2$ and $\sigma_{(1)}^2$ are the variances of the net benefit of t_2 and t_1 respectively (from table 1 and 2). What is clear from 6a is that as the sample becomes very large $\sqrt{V|n, n_{(2)}}$ tends to 1 and the EVSI approaches the EVPI, confirming the interpretation of the EVPI and the characterisation of the benefits of sampling. The EVSI in 7a represents the benefit of sample information to one individual patient but this information can be used to inform the treatment decision for the population of current and future patients.

$$7b \quad EVSI|n, n_{(2)} \cdot \sum_{t=1}^T \frac{(1 - r)^t}{(1 + r)^t}$$

The population of current and future patients is also the population from which the sample will be drawn. Those who participate in the trial will not be able to benefit from the information generated by the research because they will have already been treated¹¹. Increasing the sample size provides more information for future patients who do not participate in the trial but it also “uses up” those who would otherwise be able to benefit from the sample information (an opportunity cost of increasing sample size). We are concerned with the benefits of information to the population of current and future patients and do not explicitly model the expected costs and benefits to individual trial entrants. This issue of “individual ethical concerns”⁴⁵ is appropriately left for those who are responsible for the ethical approval of trials, however this approach will allow us to establish the opportunity cost to “collective ethics” of holding particular ethical concerns for trial entrants (if ethical approval is given and informed patients agree to participate in the trial then presumably there is “individual equipoise”)⁴⁶⁻⁴⁸.

The marginal cost of additional sample information will include the additional treatment cost compared to current practice ($C_2 - C_1$) plus any marginal reporting costs (C_r). The cost of a sample (C_s) of n with $n_{(2)}$ allocated to the experimental arm and $n - n_{(2)}$ allocated to control (current practice) and excluding any fixed cost is given by 7c.

$$7c \quad C_s|n, n_{(2)} = (C_2 - C_1) \cdot n_{(2)} + C_r \cdot n$$

The difference between the expected benefits of sample information and the cost of the sample is the expected net benefit of sample information (ENBS|n, $n_{(2)}$).

$$7d \quad \text{ENBS}|n, n_{(2)} = \text{EVSI}|n, n_{(2)} - C_s|n, n_{(2)}$$

The technically efficient scale of this research will be where the ENBS in 7d is positive and reaches a maximum (fixed cost have no effect on the optimal scale). At this point sample size is optimal and if the ENBS exceeds any fixed cost then it will be efficient to conduct the research at this technically efficient scale. This is the necessary and sufficient condition for deciding to acquire more information to inform the treatment decision. The ENBS, the EVSI and the C_s are illustrated in figures 4a and 4b for example 2 and example 1 respectively. It is not worth acquiring more sample information in both cases ($\text{ENBS}|n, n_{(2)} > 0 \forall n$) and the treatment decision for current and future patients should be based on the prior (original sample) mean. However it is also clear that the benefit of sample information is considerably higher for example 2 where the incremental net benefit was not regarded as significant, (£194,949 when $1/g = £700$ and $n = 500$) than in example 1 where it was (£4,136 when $1/g = £400$ and $n = 500$). Once again this demonstrates that difference between example 1 and 2 is not the decision about which treatment to adopt but the value of acquiring more information to inform this decision in the future.

The Expected Net Benefits of Sample Information (optimal sample allocation)

This approach has imposed a fixed and equal allocation of the sample between the two arms of the trial²⁰. This is an entirely arbitrary rule and takes no account of the marginal benefit or the marginal cost of allocating a trial entrant to the control or experimental arm. The same numerical examples demonstrate that if arbitrary fixed allocation rules are used then the research design will be technically inefficient, the value of sample information will be underestimated and there will be a danger that cost-effective research proposals may be rejected.

The equal allocation of patients between experimental and control arms of a trial is often used and is implicitly justified by assuming that the variance of the outcome of interest for the control arm of the trial is the same as the experimental arm, so that the marginal benefits (reduction in sample variance) of assigning an additional trial entrant to either arm of the trial will be the same^{31,49}. However there is little justification for this rule of precedent and whether a trial entrant should be allocated to a particular arm of a trial should be determined by the marginal benefits of assigning the patient to that arm (the variance of the net benefits of that arm) and the marginal costs of assigning the patient to that arm (the additional treatment costs).

There is a body of literature which considers the optimal allocation of trial entrants in sequential clinical trials where the results of the trial accumulate over time and can be used to assign entrants to the different arms³³. An example of this type of approach is Bather's "play the winner rule" where patients are assigned to the arm of the trial which appears to be most effective given the accumulated trial results^{50,51}. This approach and others addressing the same problem^{52,54} do not consider the marginal cost of sampling and tend to focus on minimising the potential health cost to individuals enrolled in the trial by establishing allocation rules which minimise the number of individuals enrolled in the less effective arm to achieve a specified power and level of significance. They are not concerned with the explicit valuation of sample information for current and future patients^{55,56} and are primarily concerned with sequential clinical trials where the accumulated results from earlier participants are available and are used to allocate those entering the trial. Here we address a more fundamental problem of efficient allocation in a fixed sample design where sample information is only available at the end of the trial so the value of sample information, optimal sample size and the allocation of patients must be established before any sample information is available.

To establish the optimal sample allocation an estimate of the $\text{ENBS}|n, n_{(2)}$ is required for every feasible allocation of each sample considered. In this example the variance of the control and experimental arms are equal so the marginal benefits of allocating a trial entrant to either t_2 or t_1 will be the same. However the marginal cost of allocating an entrant to the experimental arm ($C_{m(2)}$) includes the incremental cost of the new treatment and the marginal reporting cost

$$8a \quad C_{m(2)} = (C_2 - C_1) \cdot n_{(2)} + C_r \cdot n_{(2)}$$

This will be higher than the marginal cost of allocating the entrant to control $C_{m(1)}$ or current practice (assuming $(C_2 -$

$C_1 > 0$) where there will be no additional treatment costs.

$$8b \quad C_{m(1)} = C_r \cdot (n - n_{(2)})$$

In these circumstances a smaller proportion of the sample will be allocated to the experimental arm which will reduce (increase) the marginal benefits of allocating an entrant to the control arm (experimental arm). The optimal allocation will be where the marginal net benefit is equalised across both arms of the trial or where the $ENBS|n, n_{(2)}$ reaches a maximum ($ENBS|n, n_{(2)}^*$) for each sample size. This optimal contingent allocation is illustrated in table 3 (for example 2) where the possible sample sizes considered are represented by each row, the feasible allocation to t_2 are represented by the columns and the payoffs ($ENBS|n, n_{(2)}$) are in the main body of the table. The optimal allocation of each sample is the row maximum, and the optimal contingent allocations ($n_{(2)}^*$) and associated payoff ($ENBS|n, n_{(2)}^*$) are illustrated in the right hand columns.

The optimal contingent allocation of the sample is illustrated for example 2 in figure 5a. The higher marginal cost of allocating the sample to t_2 means that the optimal allocation to the experimental arm is always less than with an equal allocation rule. Indeed when the sample size is less than 380 no sample is allocated to the experimental arm because the $ENBS|n, n_{(2)} < 0 \forall n_{(2)}$. However when $n > 380$ a proportion of the sample is allocated to t_2 . This proportion falls with sample size because as the sample size increases the marginal benefit of sample information falls and any differences in the marginal benefit of assigning an entrant to either t_2 or t_1 (due to in this case to unequal allocation) will become less significant but the differences in the marginal cost of sampling remains constant. This means that the optimal allocation will change with sample size and a smaller proportion will be allocated to the more costly arm of the trial as sample size increases.

Once the $ENBS$ for each sample size (given that it will be allocated optimally between t_2 and t_1) is established, the optimal sample size will be where the $ENBS|n, n_{(2)}^*$ reaches a maximum. If this is positive and greater than the fixed cost of the research then it will be efficient to acquire more sample information to inform this treatment decision. In example 1 where the value of sample information was lower (see figure 3 and figure 4b) $n_{(2)}^* = 0$, the $ENBS|n, n_{(2)}^* < 0 \forall n$ and it is not efficient to gather more information to inform the treatment decision. In example 2 when the fixed allocation rule was imposed the $ENBS < 0$ (see figure 4a) but once an optimal sample allocation is adopted the optimal sample size is $n^* = 1,050$ with $n_{(2)}^* = 161$ allocated to the experimental arm of the trial (see figure 5b). The $ENBS|n^*, n_{(2)}^*$ at this technically efficient scale of research is £22,812 and it will now be efficient to acquire this sample information if this exceeds the fixed costs of the research. Once this research has been conducted then the treatment decision can be revised in the light of the new sample information but should still be based on the posterior mean net benefit which is a weighted average of the prior and sample means with the weights representing the informational content of each.

$$9 \quad \eta_1 = \frac{I_0 \cdot \eta_0 + I_n \cdot \eta_n}{I_0 + I_n} \quad \text{where} \quad I_0 = \frac{1}{\sigma_0^2}, \quad I_n = \frac{1}{\sigma_n^2}, \quad \sigma_n^2 = \frac{\sigma_{(2)}^2}{n_{(2)}} + \frac{\sigma_{(1)}^2}{n - n_{(2)}}$$

In this example the prior mean is 0.5714 and the prior decision is to choose t_2 . Following the proposed trial a sample mean incremental net benefit for t_2 of less than -0.4192 (the critical value of the sample mean (η_n^*))₁₂ would need to be observed before this prior decision should be reversed because this would generate a posterior mean of less than zero. Once the treatment decision has been revised in this way decision makers must consider whether more sample information should be acquired if circumstances change (for example if the budget for service provision and $1/g$ increases). The prior variance of the mean net benefit for this future decision will be the posterior variance (σ_1^2) following this trial, which is a combination of the initial prior and sample variance (and will be some fraction of both)₁₃.

Sample information will take some time to acquire and the time required will depend on the size of the sample relative to the eligible population. In this example $n^* = 1,050$ which suggests that the trial would take over 6 months even if all patients were enrolled in the trial as they presented (assuming constant rate over presentations over the year). If not all eligible patients participate then the treatment decision while the trial is being conducted should be based on the prior net benefits, which in this case indicates that the treatment should be changed from t_1 which was current practice to the new treatment t_2 . This decision may be reversed when the results of the new trial are available because there is a chance that the sample mean will be less than its critical value and the treatment decision will be switched back from t_2 to t_1 . If there are costs associated with treatment t_2 which will become sunk if

treatment is switched to t_2 then there will be a reluctance to base the current treatment decision on the prior mean when it is also decided to collect more sample information¹⁴. These costs can be incorporated in the decision rule explicitly by establishing the probability that the prior decision will change following the trial results. In this example it will be the probability that the sample mean from the trial will be less than its critical value ($p(\eta_n < \eta_n^*)$).

$$10a \quad H_0: \eta_n = \eta_0, \quad p(\eta_n < \eta_n^*) = p\left[Z > \frac{\eta_n - \eta_n^*}{\sigma_n}\right], \quad \sigma_n = \sqrt{\frac{\sigma_{(2)}^2}{n_{(2)}} + \frac{\sigma_{(1)}^2}{n^* - n_{(2)}}}$$

Since this treatment choice must be made before any other sample information is available the null is that the sample mean will be centred in the prior mean with a sample variance of (σ_n^2) . Given an estimate of the costs which will be sunk if the treatment decision is switched from current practice to the new treatment ($S_{(2)}$) the prior treatment decision will now be to choose t_2 if the expected benefits to the population of patients not enrolled in the trial $(I_t - n_t)$, for the duration of the trial $(S \leq T)$, is greater than the expected costs:

$$10b \quad \eta_0 \cdot \sum_{t=1}^{S \leq T} \frac{(I_t - n_t)}{(1+r)^t} - p(\eta_n < \eta_n^*) \cdot g \cdot S_{(2)} > 0 \quad \vee \quad S_{(2)} < \eta_0 \cdot \sum_{t=1}^{S \leq T} \frac{(I_t - n_t)}{(1+r)^t} \cdot \frac{1}{p(\eta_n < \eta_n^*)} \cdot \frac{1}{g}$$

In this numerical example $p(\eta_n < \eta_n^*) = 0.009123$ and assuming that the trial would take one year with 950 patients not participating the maximum costs which could be sunk in switching from t_1 to t_2 is approximately £41.6 million. The total treatment cost for these patients (and the maximum possible sunk cost) will be £1.9 million therefore the treatment decision should be to switch from t_1 to t_2 while the trial is being conducted. The decision may be revised and switch back from t_2 to t_1 if the sample (posterior) mean is less than the critical value (zero) when this trial is completed.

These numerical examples have demonstrated that adopting arbitrary sample allocation rules even in a fixed sample design will lead to a technically inefficient scale of research; the expected net benefit of sample information will be underestimated; and there will be a danger that cost-effective research will be rejected. These examples also demonstrate once again that inference is irrelevant to the treatment decision. The difference between example 1 and 2 is not the treatment decision but the decision to acquire more information. The distribution of the incremental net benefits is only relevant to the decision of whether more information should be acquired to inform treatment choice: a decision which should be based only on the posterior mean net benefit.

5 Discussion

The approach outlined above demonstrates that a decision making approach to the stochastic evaluation of health care technologies is both necessary and feasible. Decisions should be based only on the mean net benefit irrespective of whether differences are statistically significant or fall outside a Bayesian range of equivalence. The distribution of the incremental net benefits is only relevant in deciding whether it is worth collecting more information about the decision problem to reduce the expected costs of basing decisions on the existing estimates of the mean net benefit. This framework for decision making can distinguish the simultaneous but conceptually separate steps of deciding which alternatives should be chosen, given existing information, from the question of whether more information should be acquired. It mirrors the sequential nature of decision making: making an initial decision; deciding to gather evidence; revising decisions in the light of this new information; and again considering whether more information is required. It also ensures that the type of information acquired (research) is driven by the objectives of the health care system and is valued in a way which is consistent with the budget constraint on service provision. Indeed inference, whether based on Bayesian ranges of equivalence or on classical frequentist tail area probabilities, is redundant because it is inconsistent with the objectives of any coherent health care system.

This approach to the value of information means that clinical research can be designed and research priorities can be set in a way which is consistent with the objectives of service provision and the budget constraint faced by clinicians. Although treatment decisions should be based only on the mean net benefit, irrespective of their distribution, the approach does not necessarily imply that small clinical trials will be acceptable or that the treatment decisions can be based on poor quality evidence. Indeed in some circumstances (where the prior mean incremental net benefit is close to zero with large variance, the marginal costs of sampling is low, the value of $1/g$ is high and the incidence of

patients entering the decision problems is also high) very large trials, as large or possibly even larger than those demanded by Collins, Peto, Gray, Parish (1996) ⁵⁸, may well be efficient. The point is that the amount of information which should be acquired (whether through clinical trials or econometric studies of observational data which adjust for selection bias ^{59, 60}) is an empirical rather than an ideological question. The technically efficient scale of research will differ across different clinical decision problems, different budgets and alternative judgements about the objective of a health care system.

This also has implications for the regulation of new health care technologies. Clearly a regulatory framework which demands arbitrary clinical improvements at 5% significance and 80% power imposes substantial opportunity costs. Indeed the regulation of new pharmaceuticals may well demand evidence which is not necessarily relevant to the objectives of the health care system: distorting the licencing of new technologies and research and development priorities. A rational approach to regulation based on an assessment the value of additional information would be consistent with the objectives and the budget for service provision. Efficient regulation would demand more information for some new technologies as compared to others. Similarly efficient regulation would require different information for the same new technology under different budgetary regimes. This is not to argue that regulation and the quality of evidence should be relaxed but to suggest that regulators should ask if it is efficient to gather more information about a new technology before it is licenced. In some circumstances the value of information and the technically efficient scale of research will be relatively large (outlined above) and much more information than currently demanded may be required. However in other circumstances, such as a new technology with good prior evidence of substantial net benefits and a relatively small population of eligible patients, the value of information will be lower and the efficient scale of research will also be lower. The current system of demanding the same evidence (often only about efficacy rather than effectiveness and efficiency) irrespective of the issues discussed above clearly cannot be efficient and the consequences of arbitrary regulation will be distorted research and development priorities and distorted licencing decisions which will impose costs on current and future patients.

However whatever view is taken about these policy issues when it comes to informing decisions which cannot be deferred classical frequentist statistical inference is entirely bankrupt and only remains standing by virtue of its longevity. Surely it is time to adopt an approach which addresses the decisions which must be made and which economic or indeed clinical evaluation claims to inform. The opportunity costs of continuing to make decisions based on the arcane and arbitrary rules of inference are very real and can be measured in terms of resources or life years forgone.

Notes

- 1 It has been suggested that average cost effectiveness ratios can be used to compare non mutually exclusive programmes, however the appropriate ratios to compare across programmes are the relevant incremental ratios generated by comparing the mutually exclusive options within each programme. If this involves a comparison with a do nothing option then this is still a incremental ratio where one alternative happens to have zero costs but not necessarily zero expected health outcome.
- 2 Without this information neither CEA or CUA can address allocative efficiency and can only make statements about technical efficiency in the case of dominance. The view that CEA can only address technical efficiency but CUA can address allocative efficiency is clearly false: once the shadow price of the budget constraint has been established both can address allocative efficiency. The only difference between the two is in terms of the dimensions of health outcome deemed relevant (a normative judgement of a social decision maker in the social decision making or extra-welfarist approach). Establishing a value of g is particularly important in sequential clinical decision problems. For example in a test-treatment decision problem a contingent treatment decision must be made before the incremental cost-effectiveness ratio for the initial diagnostic decision can be established and the explicit monetary valuation of health outcome can not be avoided even before the evaluative study has been designed and the alternatives which are regarded as relevant have been identified
- 3 Clearly the net benefits measured in terms of health outcome will not be the same because the value placed on health outcome differs in these two situations. We could construct an example where the net benefits were the same in terms of health outcome but then the net benefits measured in money terms would not be equal.
- 4 If the budget is variable then it has been suggested that a fixed price should be adopted which may be based on the views of social decision makers or the preferences of different groups of potential patients (Johannesson and Karlsson 1997) which in certain circumstances will be equivalent to a Paretian approach which elicits individual willingness to pay (Garber and Phelps 1997). If the budget is fixed by some bureaucratic or political process then the value of g will be the shadow price of the budget constraint which given information about each health care programme currently undertaken (and for each possible scale of each programme) can be found by solving the dual in a linear programme where the objective of the primal is to maximise health gain subject to the budget constraint (Stinnett and Paltiel 1996, Stinnett and Paltiel 1997). Clearly which budget is considered fixed (whether it is the national budget for all health care or for particular patient groups, broad clinical area, specialities, geographical areas, or particular service providers or purchasers) will be determined by the perspective of the evaluation in an extra welfarist approach. In the absence of full information a second best rule is to compare the new technology to some programmes which are currently funded within the budget that is regarded as fixed and relevant to the decision makers perspective. If the CER is lower than any one of these programmes then this is a necessary and sufficient condition to establish its cost effectiveness. Finding that the CER of the new technology is greater than the CERs of the comparator programmes is a necessary but not sufficient condition for concluding that the new technology is not cost effective because a less cost-effective programme may exist and be currently funded within the budget. Also if we wish to establish whether a new technology will be cost-effectiveness in the future then some predictions about the growth in the budget will be required as well as information on the expected cost inflation of existing programmes and the cost-effectiveness of future technologies which will be competing for funding. Finally the value of g will be specific to the measure of health outcome that is deemed to be appropriate, whether this is single dimensional clinical measure of effectiveness, an index of health related quality of life, or some measure of utility. These observations simply reinforce the view that this crucial parameter should be explicitly established and not left to the intuitive judgement of decision makers.
- 5 For comparability between section 2 and 3 we assume that the original prior was uninformative in which case the sample means illustrated in figure 1 and 2 are also posterior means and the sample variance is also the posterior variances see (9) and foot notes 11 and 12.
- 6 The current regulation of pharmaceuticals places great emphasis on avoiding rare but catastrophic events for example the small relative risk of thrombosis with the third generation oral contraceptives. It suggests an overriding (above risk aversion) concern not to "do harm". These preferences for safety can be explicitly reflected in the mean net benefit if the outcome of the catastrophic event is given sufficient weight
- 7 This approach is easily generalisable to sequential trials. If the results of the trial accumulate then the treatment decision can be revised after each result and the decision to continue to collect information can also be made after each result. In a sequential trial can keep running the trial until we are satisfied to make the treatment decision based on the posterior mean. However this does not mean that no more trials would be conducted in the future if the budget for service provision changes then $1/g$ will either increase or decrease it may be worth acquiring more information to inform the decision in new circumstances
- 8 See (9) and foot notes 11 and 12 where an uninformative prior can be characterises as $I_0=0$ in 9 and contains no information in which case $\eta_1=\eta_0$ and $\sigma_1 = \sigma_0$
- 9 Since the original prior is assumed to be uninformative $\sigma_0^2 = \text{Var}(\eta)/n$ in 4a (see 9 and foot notes 11 and 12), and we assume that the original sample of 200 observations on η was allocated equally with 200 entrants assigned to the experimental arm and 200 to the control arm of the trial.
- 10 It should be noted that the value of g will determine the slope of the loss function but it will also influence σ_0 and η_0 . The intuition that if the budget is relaxed and society is willing to pay more for health outcome then the value of information will increase is not necessarily true for a particular clinical decision problem. Although if initially $\eta_0 < 0$ then the value of information will unambiguously rise with $1/g$ because the effect on the loss function and on η_0 will work in the same direction (see Claxton and Posnett 1996, and Claxton 1997).
- 11 Clearly there are circumstances such as a recurrent disease (non curative treatment) when individual patients will present for treatment at some later date following participation and treatment in a trial. In this situation the population that can benefit from the results of the clinical trial includes all the expected presentations which may include the same individuals entering the decision problem a number of times. This does not pose a problem for this approach it simply means that the population of current and future patients will be larger once repeated presentation is taken into account
- 12 Before sample information is available and when $\eta_0 > \eta_b = 0$ the clinician should select t_2 based on the prior mean, but once sample information is available the clinician should change this prior decision and select t_1 if $\eta_1 < 0$. The critical value of the sample mean (η_n^*) is the sample mean which generates a posterior mean that changes the prior decision. This can be found by setting $\eta_1=0$ and $\eta_n = \eta_n^*$ in 9 and rearranging:

$$\eta_n^* = \frac{(I_0 + I_n)\eta_b - I_0\eta_0}{I_n}$$

If the sample mean is less than this critical value ($\eta_n < \eta_n^*$) the posterior mean will be less than zero and the clinician should select t_1 , but when ($\eta_n > \eta_n^*$) the posterior mean is greater than zero and the clinician should select t_2 .

- 13 The prior variance of the mean net benefit for this future decision will be the posterior variance (σ_1^2) following this earlier trial. The posterior variance is a combination of the initial prior and sample variance (σ_n^2) and will be some fraction of both the prior and sample variance.

$$\sigma_1^2 = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_n^2} \cdot \sigma_n^2 \quad \vee \quad \sigma_1^2 = \frac{\sigma_n^2}{\sigma_n^2 + \sigma_0^2} \cdot \sigma_0^2$$

- 14 If there are no sunk costs associated with switching between treatments then changes in treatment choice over time would simply reflect that fact that treatment decisions are based on the best available evidence of efficiency at that time. Clearly as the evidence accumulates the efficient treatment choice may well change as the best available evidence changes.

References

- 1 O'Brien, BJ. Drummond, MF. Labelle, RJ. et al. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 1994; 32: 150-63
- 2 O'Brien BJ. and Drummond, MF. Statistical versus quantitative significance in the socioeconomic evaluation of medicine. *Pharmacoeconomics* 1994; 5: 389-398.
- 3 Wakker, P. and Klaassen, MP. Confidence intervals for cost/effectiveness ratios, *Health Economics* 1995; 4:373-381.
- 4 Willan, AR. and O'Brien, BJ. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Economics* 1996; 5: 297-305.
- 5 Mullahy, J. What you don't know can't hurt you? Statistical issues in designing medical technology evaluation standards, *Medical Care* 1996; 34: 124-135.
- 6 Mullahy, J. Which cost-effectiveness ratio? Evaluation of health policies and medical technologies in a stochastic context. Unpublished draft 1997.
- 7 Garber, AM., Weinstein, MC. Torrence, GW. and Kanlet, MS. Theoretical foundations of cost-effectiveness analysis, In: Gold, MR., Siegal JE., Russell, LB. and Weinstein, MC. eds. *Cost-effectiveness in health and medicine*. New York, Oxford University Press, 1996.
- 8 Weinstein, MC. Principles of cost-effective resource allocation in health care organisations. *International Journal of Technology Assessment in Health Care*. 1990; 6: 93-103
- 9 Weinstein, MC. From cost-effectiveness ratios to resource allocation: where to draw the line? In: Sloan, FA. editor. *Valuing health care: cost, benefits and effectiveness of pharmaceuticals and other medical technologies*. Cambridge: Cambridge University Press, 1995.
- 10 Johannesson, M. and Weinstein, MC. On the decision rules of cost-effectiveness analysis. *Journal of Health Economics* 1993; 12: 459-467
- 11 Weinstein, MC. and Zeckhauser, RJ. Critical ratios and efficient allocation. *Journal of Public Economics* 1972; 2: 716-721.
- 12 Phelps, CE. and Mushlin, AI. On the near equivalence of cost-effectiveness and cost-benefit analysis. *International Journal of Technology Assessment in Health Care* 1991; 7: 12-21.
- 13 Phelps, CE. and Mushlin, AI. Focusing technology assessment using medical decision theory. *Medical Decision Making* 1988; 8: 279-289.
- 14 Stinnett, AA. and Paltiel, AD. Mathematical programming for the efficient allocation of health care resources. *Journal of Health Economics* 1996; 15: 641-653.
- 15 Stinnett, AA. and Paltiel, AD. Birds of a feather or flamingoes and mustard? Fixed budgets and fixed prices for health care resource allocation. *Journal of Health Economics* 1997; forthcoming
- 16 Johannesson, M. and Karlsson, G. On the efficient allocation of health care resources, *Journal of Health Economics* 1997; forthcoming.
- 17 Claxton, K. The value of clinical information: an economic approach to research priority setting. University of York, *D.Phil. Economics* 1997.
- 18 Claxton, K. and Posnett J. The role of decision analysis in structuring the economic evaluation of sequential clinical decision problems. University of York: *Discussion Papers in Economics* No 96/24, 1996.
- 19 Garber, AM. and Phelps, CE. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* 1997; 16: 1-31.
- 20 Claxton, K. and Posnett J. An economic approach to clinical trial design and research priority setting. *Health Economics* 1996; 5: 513-524.
- 21 Stinnett, AA. and Mullahy, J. Net Health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Unpublished draft 1997.
- 22 Culyer, AJ. The normative economics of health care finance and provision. In: McGuire, A. Fenn, P. and Mayhew, K. editors. *Providing health care: the economics of alternative systems of finance and provision*. Oxford: Oxford University Press, 1991.
- 23 Culyer, AJ. and Evans, RG. Mark Pauly on welfare economics: normative rabbits from positive hats. *Journal of Health Economics* 1996; 15: 243-251.
- 24 Reinhardt, UE. Reflections on the meaning of efficiency: can efficiency be separated from equity? *Yale Law and Policy Review* 1992; 10: 302-315.
- 25 Pauly, M. Editorial: A re-examination of the meaning and importance of supplier-induced demand, *Journal of Health Economics* 1994; 13(3): 369-372.
- 26 Schwartz, D. and Lellouch, J. Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Disease* 1967; 20: 637-648.
- 27 Spiegelhalter, DJ, Freedman, LS and Parmar, MKB. Bayesian approaches to randomised trials. *Journal of the Royal Statistical Society A* 1994; 157: 357-416.
- 28 Lindley, DV. Discussion of the paper by Spiegelhalter, Freedman and Parmar. *Journal of the Royal Statistical Society A* 1994; 157: 393.
- 29 Berry, DA. Discussion of the paper by Spiegelhalter, Freedman and Parmar. *Journal of the Royal Statistical Society A* 1994; 157: 399.
- 30 Whitehead, J. *The Design and Analysis of Sequential Trials*. Ellis Horwood, 1992.
- 31 Lachin, JM. Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* 1981; 1: 13-28.

- 32 Spiegelhalter, DJ. and Freedman, LS. Bayesian approaches to clinical trials. *Bayesian Statistics Volume 3*. Oxford: 1988: 453-477.
- 33 Armitage, P. The search for optimality in clinical trials. *International Statistical Review* 1985; **53**:15-24.
- 34 Armitage, P. *Sequential Medical Trials*. Oxford: Blackwell, 1975.
- 35 Anscombe, FJ. Sequential medical trials. *Journal of the American Statistical Association* 1963; **58**: 365-383.
- 36 Colton, T. A model for selecting one of two medical treatments. *Journal of the American Statistical Association* 1963; **58**: 388-400.
- 37 Colton, T. A two stage model for selecting one of two treatments. *Biometrics* 1965; **21**: 169-180.
- 38 Deber, R. and Goel V. Using explicit decision rules to manage issues of justice, risk and ethics in decision analysis: when is it not rational to maximize expected utility? *Medical Decision Making* 1990; **10**: 181-194.
- 39 Hey, JD. Decisions under uncertainty, In: van der Ploeg F eds. *Mathematical method in economics*. John Willey and Sons 1984.
- 40 Loomes, G. And Sugden, R. Regret theory: an alternative theory of rational choice under uncertainty, *Economics Journal* 1982; **92**: 805-824.
- 41 Machina, MJ. Choice under uncertainty: problems solved and unsolved, *Economic Perspectives* 1987; **1**: 121-154.
- 42 Schlaifer, R. *Introduction to Statistics for Business Decisions*. McGraw-Hill, 1961.
- 43 Raiffa, H. and Schlaifer, R. *Probability and Statistics for Business Decisions*. McGraw-Hill, 1959.
- 44 Raiffa, H. *Decision analysis: introductory lectures on choices under uncertainty*. Addison-Wesley, 1968.
- 45 Pocock, SJ. When to stop a clinical trial. *British Medical Journal* 1992; **305**: 235-240.
- 46 Lilford, R. and Jackson, J. Equipoise and the ethics of randomization. *Journal of the Royal Society of Medicine* 1995; **88**: 552-559.
- 47 Johnson, N. Lilford, RJ. and Brazier, W. At what level of collective equipoise does a clinical trial become ethical. *Journal of Medical Ethics* 1991; **17**: 23-35
- 48 Freedman, B. Equipoise and the ethics of clinical research. *New England Journal of Medicine* 1987; **317**: 141-145.
- 49 Schwartz, D. Flamant, R. and Lellouch, J. *Clinical Trials*. London, Harcourt Brace Jovanovich 1980
- 50 Bather, JA. Randomised allocation of treatments in sequential experiments. *Journal of the Royal Statistical Society B* 1981; **43**: 265-292.
- 51 Bather, JA. On the allocation of treatment in sequential medical trials. *International Statistical Review* 1985; **53**: 1-13.
- 52 Hoel, DG. Sidel, M. and Weiss, GH. A survey of adaptive sampling for clinical trials. In: Elashoff, R. editor. *Perspectives in Biometry*. New York: Academic Press, 1975.
- 53 Igiewicz, B. Alternative designs: sequential, multi-stage, decision theory and adaptive designs. In: Buyse, ME. Staquet, J. and Sylvester, RJ. editors. *Cancer clinical trials: methods and practice*. Oxford University Press, 1983.
- 54 Zelen, M. Play the winner rule and the controlled clinical trial. *Journal of the American Statistical Association* 1969; **64**: 131-146.
- 55 Berry, DA. A Bernoulli two-armed bandit. *Annals of Mathematical Statistics* 1972; **43**: 871-897.
- 56 Berry, DA. Modified two-armed bandit strategies for certain clinical trials. *Journal of the American Statistical Association* 1978; **73**: 339-345.
- 57 Upton, GJG. and Lee, RD. The importance of the patient horizon in the sequential analysis of binomial clinical trials. *Biometrika* 1976; **63**: 335-342.
- 58 Collins, R., Peto, R., Gray, R. and Parish, S. Large scale randomised evidence: trials and overviews. In: Weatherall, D., Ledingham, JGG. and Warrell, DA. Eds. *Oxford Textbook of Medicine*. Vol 1. Oxford: Oxford University Press, 1996.
- 59 Manning, R. and Claxton, K. Experimental and econometric solutions to selection bias. *Paper presented at the Health Economists Study Group*, July 1997, York.
- 60 McCellan, M. and Newhouse, JP. The marginal cost-effectiveness of medical technology: a panel instrumental-variables approach. *Journal of Econometrics* 1997: 39-64.

Figure 1a Incremental Net Benefit and Confidence Limits (Example 1)

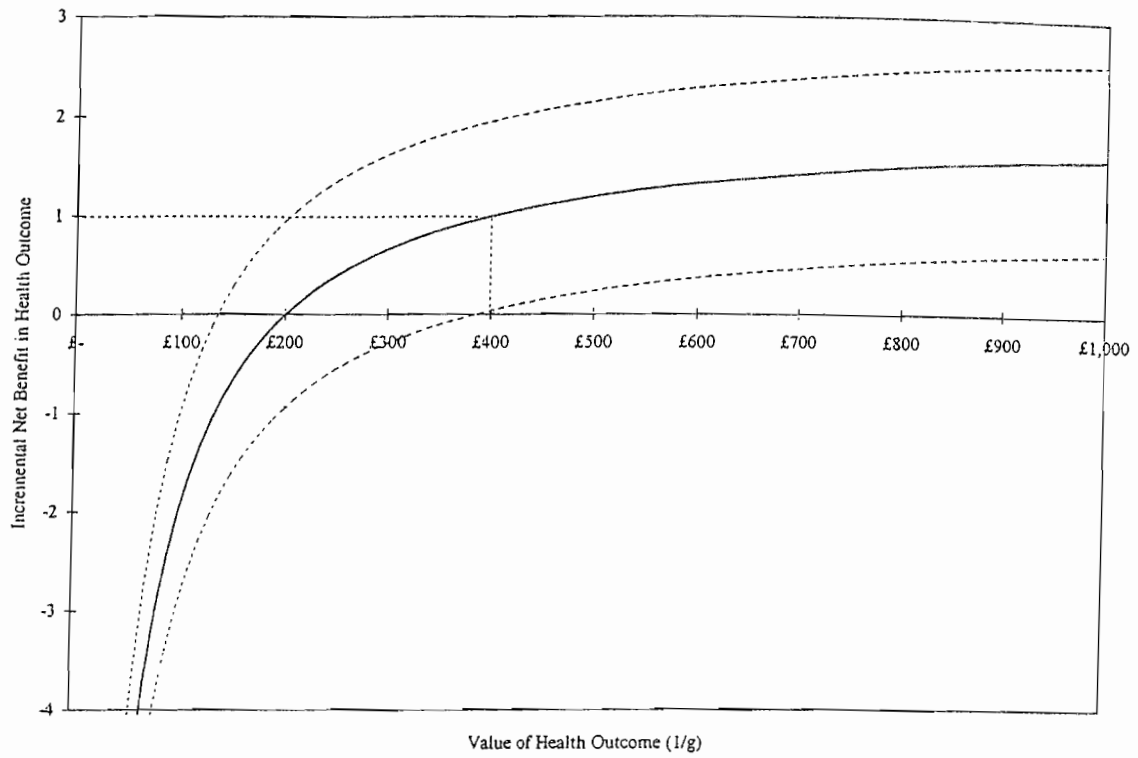


Figure 1b Incremental Net Benefit and Confidence Limits (Example 1)

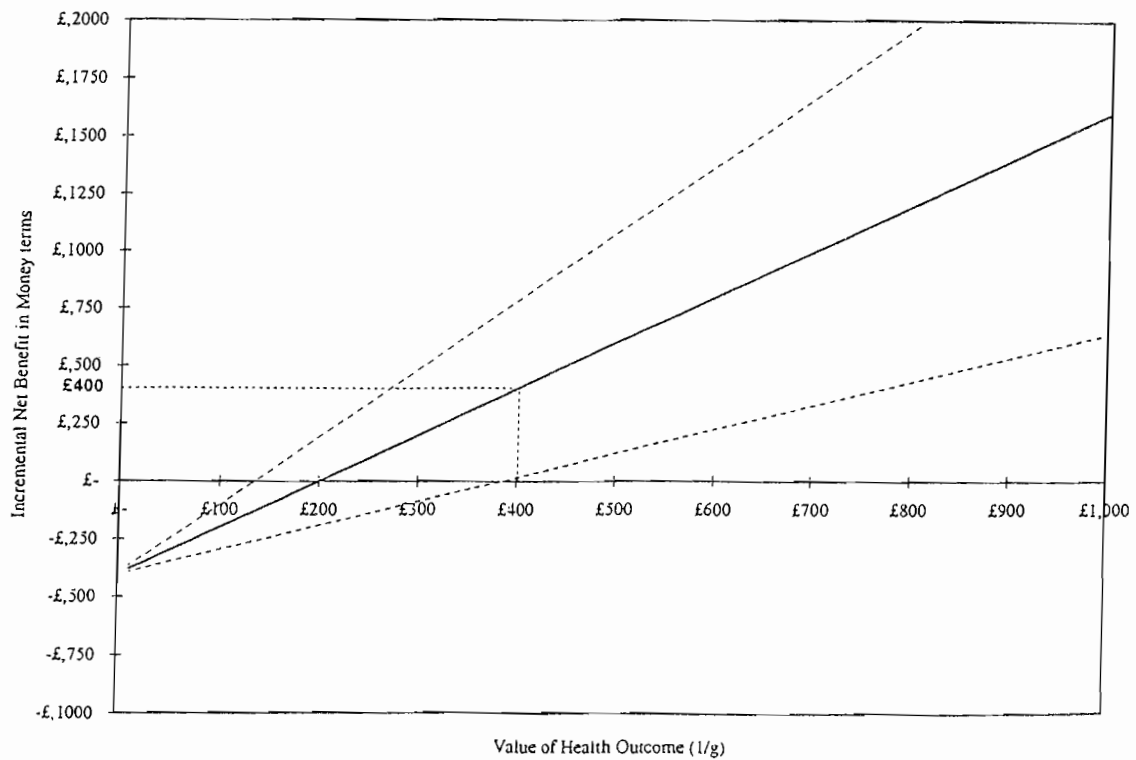


Figure 2 Incremental Net Benefit and Confidence Linimits (Example 2)

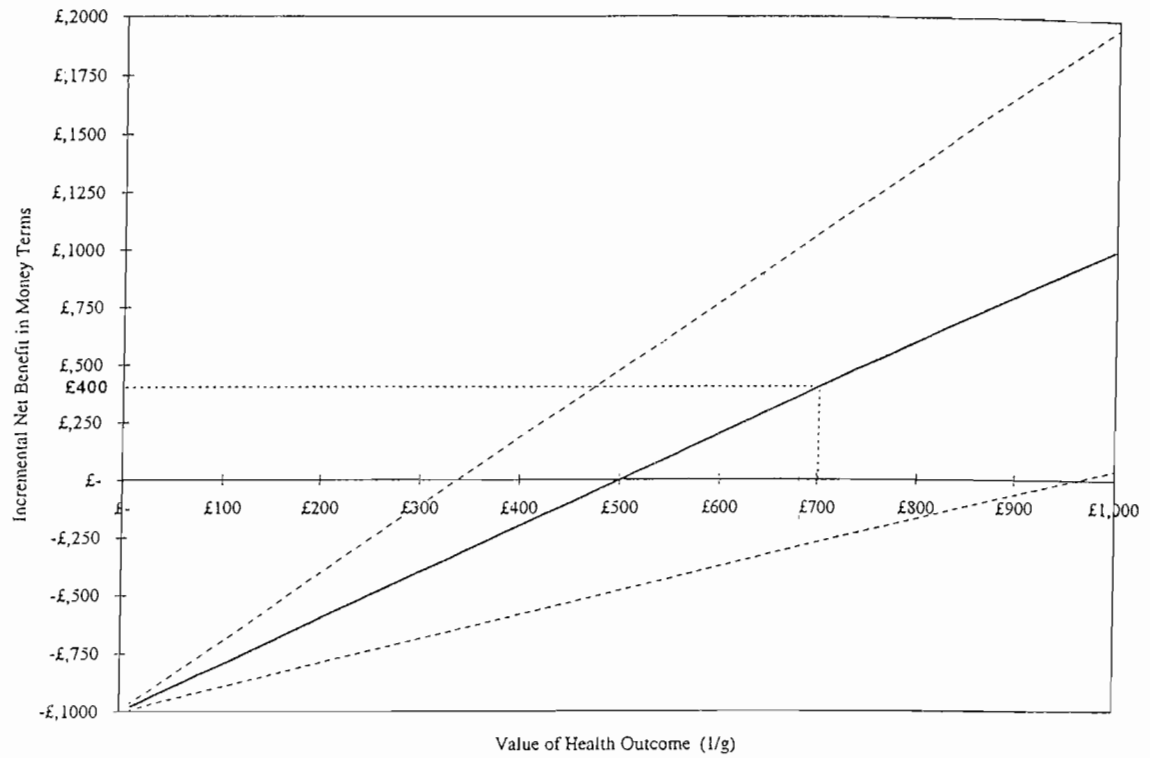


Figure 3 The Expected Value of Perfect Information (Examples 1, and 2)

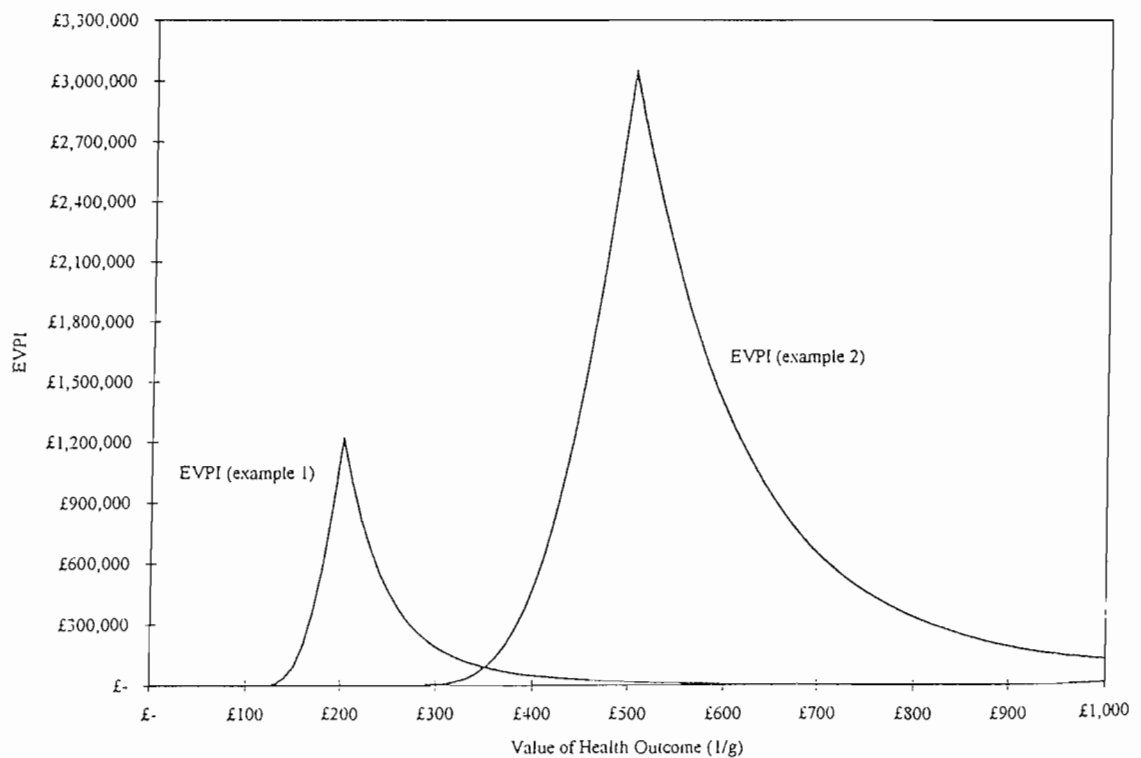


Figure 4a ENBS, EVSI and Cs with Equal Sample Allocation (Example 2)

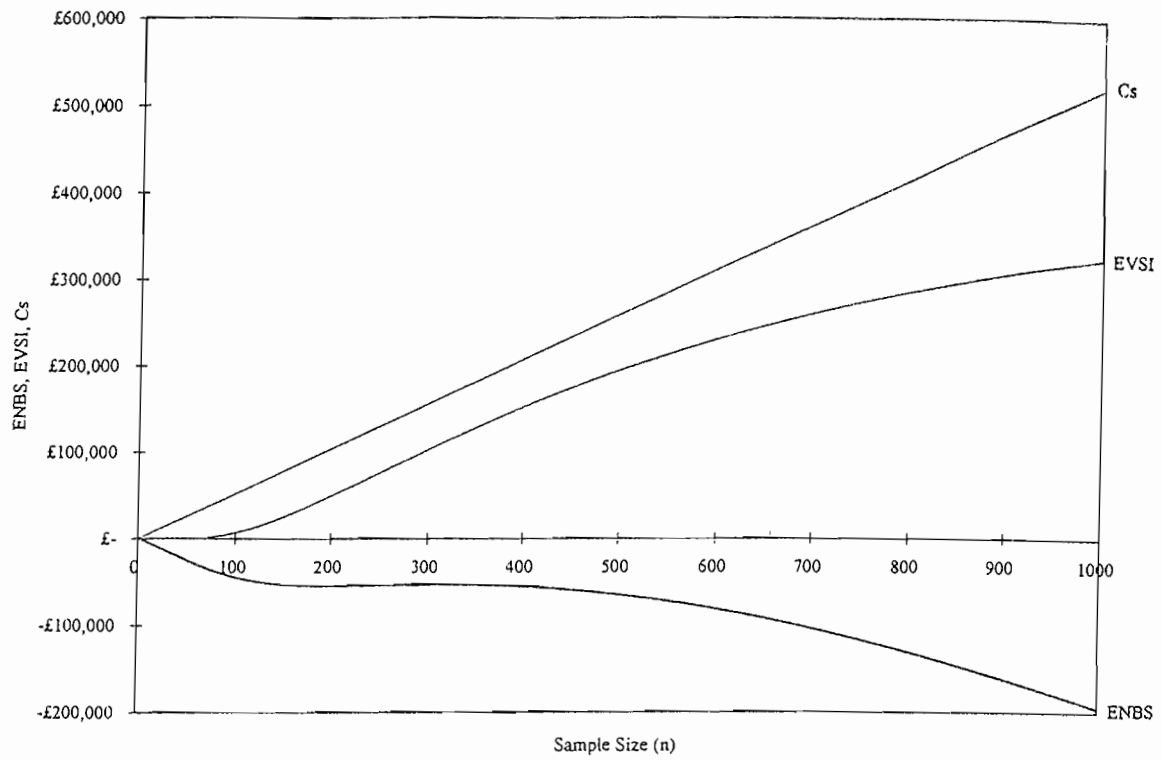


Figure 4b ENBS, EVSI and Cs with Equal Sample Allocation (Example 1)

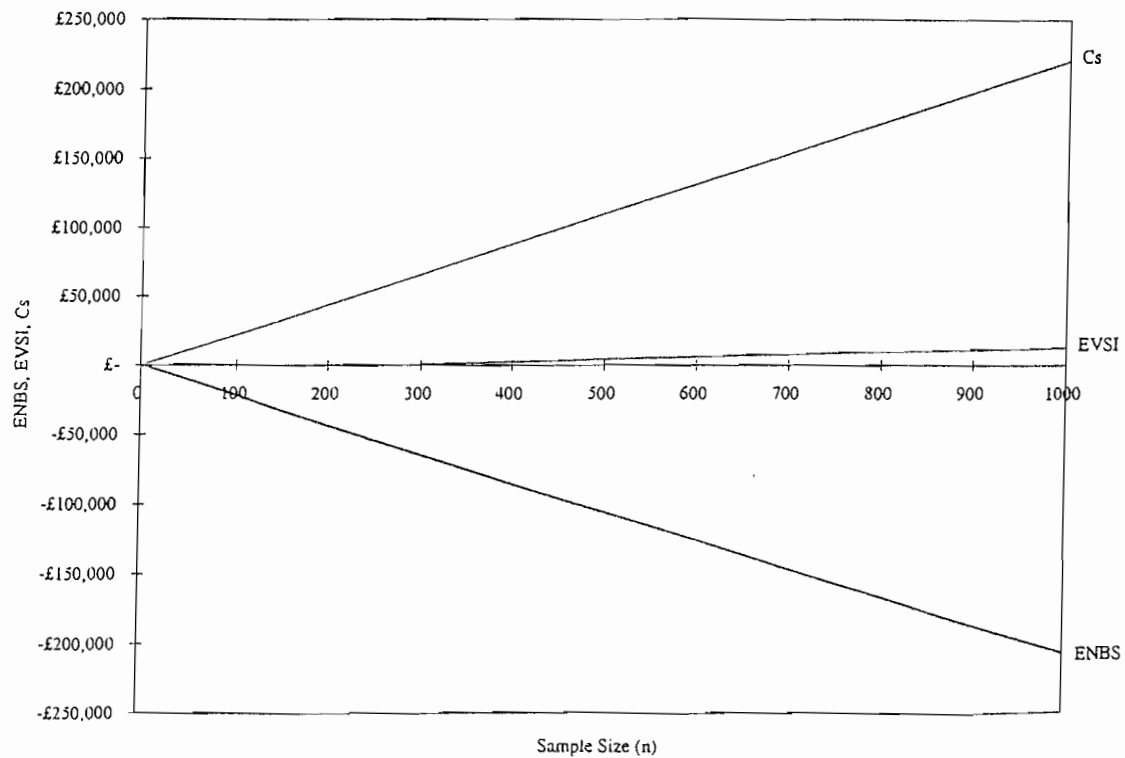


Figure 5a Optimal Sample Allocation to Treatment t_2 (Example 2)

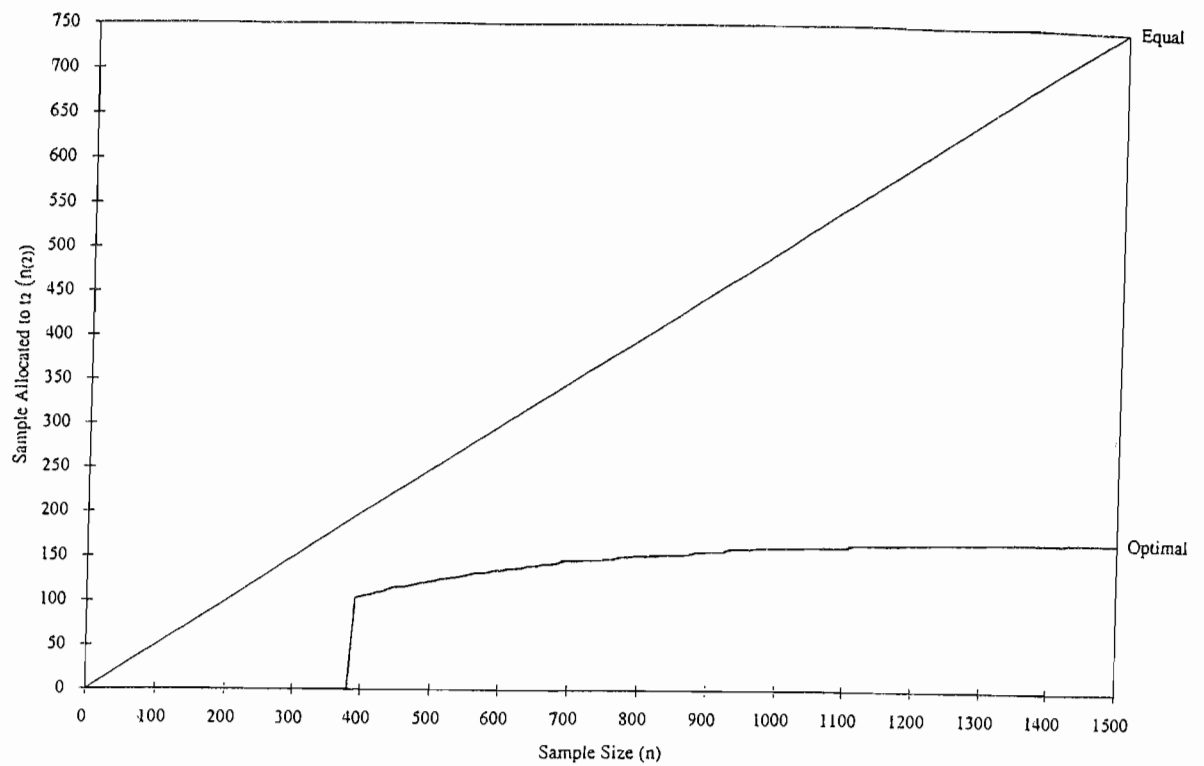
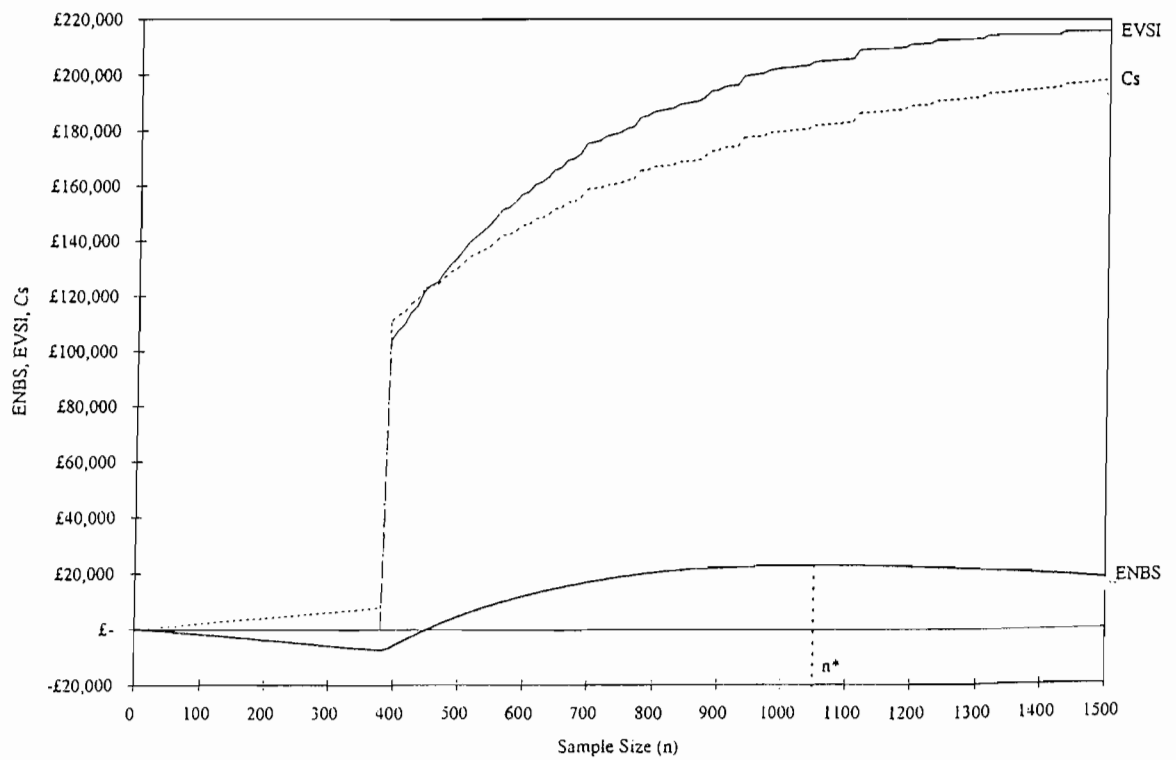


Figure 5b ENBS, EVSI and C_s with Optimal Sample Allocation (Example 2)



Appendix A

Table 1 Example 1 ($1/g=\pounds 400$, $n=200$, $\sigma_{ce}=0$)

	Sample Mean	Variance	Sample Variance
U_2	10	24	0.12
U_1	8	24	0.12
$U_2 - U_1$	2	48	0.24
C_2	£1,400	3000	15
C_1	£1,000	3000	15
$C_2 - C_1$	£400	6000	30
η	1	48.0375	0.2401875
μ	£400	7686000	38430
ICER	£200		

Table 2 Example 2 ($1/g=\pounds 700$, $n=200$, $\sigma_{ee}=0$)

	Sample Mean	Variance	Sample Variance
U_2	10	24	0.12
U_1	8	24	0.12
$U_2 - U_1$	2	48	0.24
C_2	£2,000	3000	15
C_1	£1,000	3000	15
$C_2 - C_1$	£400	6000	30
η	0.571429	48.0375	0.2401875
μ	£400	7686000	38430
ICER	£500		

Table 3 Optimal Contingent Allocation (Example 2, $1/g=\pounds 700$)[illegible]