Are head-to-head trials of biologics needed? The role of value of information methods in arthritis research

Nicky J. Welton¹, Jason Madan¹ and Anthony E. Ades¹

Abstract

Reimbursement decisions are typically based on cost-effectiveness analyses. While a cost-effectiveness analysis can identify the optimum strategy, there is usually some degree of uncertainty around this decision. Sources of uncertainty include statistical sampling error in treatment efficacy measures, underlying baseline risk, utility measures and costs, as well as uncertainty in the structure of the model. The optimal strategy is therefore only optimal on average, and a decision to adopt this strategy might still be the wrong decision if all uncertainty could be eliminated. This means that there is a quantifiable expected (average) loss attaching to decisions made under uncertainty, and hence a value in collecting information to reduce that uncertainty. Value of information (VOI) analyses can be used to provide guidance on whether more research would be cost-effective, which particular model inputs (parameters) have the most bearing on decision uncertainty, and can also help with the design and sample size of further research. Here, we introduce the key concepts in VOI analyses, and highlight the inputs required to calculate it. The adoption of the new biologic treatments for RA and PsA tends to be based on placebo-controlled trials. We discuss the possible role of VOI analyses in deciding whether head-to-head comparisons of the biologic therapies should be carried out, illustrating with examples from other fields. We emphasize the need for a model of the natural history of RA and PsA, which reflects a consensus view.

Key words: Value of information, Expected value of perfect information, Biologic therapies.

Introduction

Most of the evidence on the biologic therapies has come from placebo-controlled trials. Indirect comparisons have been carried out but they do not provide decisive information on the relative effectiveness of the biologics against each other [1]. It seems, at first, obvious that head-to-head trials are needed.

Decisions about commissioning or funding treatment trials are usually taken by committees of experts. The current state of knowledge, the efficacy of existing treatments and the size of the patient population that will benefit will naturally be taken into account. It would be relatively rare, however, for a decision to undertake a trial to be influenced by the cost of the treatment, or its cost-effectiveness, even in countries like the UK, where a recommendation to use a drug in the National Health Service (NHS) depends above all on cost-effectiveness. Value of Information (VOI) analysis [2] bridges the gap between the decision whether or not to adopt a treatment and the decision on whether or not to carry out research: it makes the resource put towards research commensurate with the health benefits that are expected to result from that research.

Figure 1 illustrates what might be termed the research cycle [3]. A decision as to which treatment to use starts with a systematic review and synthesis of the current evidence. The synthesis is embedded into a cost-effectiveness analysis (CEA), which tells us which treatment represents the best value for money. The same CEA is then used in VOI analysis. The VOI analysis can determine whether it is cost-effective to carry out further research to reduce uncertainty in the CEA model inputs (parameters); also which parameters should be researched, and even which designs and study size. Once this research is complete, we cycle again, updating the decision in light of the new evidence. In the remainder of this article, we will describe the different kinds of VOI calculation. We show what the calculations are doing in conceptual terms, but sparing readers the mathematical details, and illustrating the ideas with some examples.
Fig. 1 Six stages of the research cycle for health-care technologies, extending the concept as envisaged by Sculpher et al. [3]. Adapted from Ref. [3].

To date the methods have rarely been applied in the area of rheumatology (but see [4]) and so the examples shown are from other fields. However, we indicate how the methods could be applied to prioritize and design new research in arthritis. In the final part of the article, we outline the lessons that must be carried out when considering the potential role of VOI calculations in designing studies of biologic therapies in RA and PsA, and some of the special circumstances surrounding the role of head-to-head trials.

The different forms of VOI analysis

VOI analysis must begin with a CEA based on a clear decision question, such as: which of the biologic therapies are cost-effective for RA patients who have failed on MTX?; what is the optimal sequence in which to offer biologic therapies for RA patients who have failed on MTX?; and which of the biologic therapies are cost-effective for RA patients with a particular biomarker present? Health gains must be given monetary values, so that benefits and costs can be measured in the same units and traded against each other. Such models have become increasingly common, and are in fact required by some national reimbursement agencies like the National Institute of Health and Clinical Excellence (NICE) in the UK. Lifetime health gains are expressed in Quality Adjusted Life Year (QALY) units [5], and then a monetary value is ascribed to the QALY—for example, £20 000 per QALY is an accepted benchmark in the UK.

A further requirement in the CEA is that the uncertainty in the model parameters is recognized, which is achieved by assigning distributions to parameters, based on the available data. These may be independent distributions where each parameter is informed by a separate item of data, or a complex joint distribution based on the simultaneous estimation of several parameters from the same data. In the absence of evidence to inform such distributions, they can be elicited from clinical experts [6, 7].

Expert elicitation of course introduces some subjectivity, but a careful choice of experts across a range of backgrounds can mitigate against this. The purpose of probabilistic CEA [8, 9] is to average over these distributions to find the average benefit (also called the expected benefit) and the average cost.

We may imagine a set of decision options facing a decision maker: perhaps a choice between no treatment and several different biologic therapies in a specific target population. The decision maker should select the treatment with the highest expected net benefit, where net benefit is monetarized health gain minus the cost. If we accept the £20 000 per QALY benchmark, for example, then the net benefit of a strategy is £20 000 \times (QALY gain of the strategy) − cost of the strategy. The costs and the gain may be relative to no treatment, in which case the net benefit of no treatment is exactly zero.

Since there is uncertainty in the model inputs (parameters), the decision maker must consider the expected net benefit, averaging over all the possible values of the uncertain parameters. The expected net benefit of all the strategies must be examined, and the strategy with the highest net benefit is chosen. Let us call this $S^*$. $S^*$ represents the best strategy given the currently available data. In a probabilistic model, one often sees cost-effectiveness acceptability curves (CEACs) [10, 11]; these plot out the probability that each strategy is the cost-effective one, i.e. the one with the highest net benefit, as a function of the monetary value of a QALY. Figure 2 illustrates a CEAC for a decision problem relating to four possible strategies to increase attendance for breast cancer screening [12]. The CEACs show that at a willingness-to-pay of £20 000 per QALY, the Letter strategy was most likely to be cost-effective ($S^* = \text{Letter}$), but only just: a decision maker willing to pay £25 000 per QALY should choose the Both strategy. Note that the probability that the optimal strategy (Letter) was the best strategy was only ~44%. In the case of biologic therapies for RA, the CEAC would tell us the probability that each of the different biologics is the most cost-effective. If there is little difference in costs and benefits, then we would expect these probabilities to be roughly equal across the different biologics. If one of the biologics is better value for money then it would have a higher probability of being cost-effective than the other biologics.

Expected value of perfect information

Although Letter is best based on current evidence (Fig. 2), there is a strong probability (56%) that we would get better value for money from another strategy. If we eliminated all the uncertainty in all of the model inputs by collecting far more data, we would know for sure which strategy was the best. The implication is clear: by taking a decision based on imperfect evidence there is a chance we are making the wrong decision, and hence that we are not getting the best value for money. There is an expected loss attaching to uncertainty in the model inputs, and we can calculate how great this expected loss is likely to be.

For each possible set of parameter values that could occur we find the new best strategy $S^{new}$ that gives the greatest net benefit for that set of parameter values. We record the difference in net benefit that arises from using $S^{new}$ instead of the old best strategy based on
current evidence, $S^*$. If the best strategy is unchanged ($S^{\text{new}} = S^*$) then there is no loss, but for parameter values where the best strategy changes, then there is a positive loss in health benefits from using $S^*$. Averaging this loss over all possible sets of parameter values that can occur (weighted by how likely they are to occur) gives us the expected loss attaching to parameter uncertainty, termed the expected value (EV) of perfect information (EVPI). EVPI is the maximum net gain per person that we can obtain from running new research. If EVPI for biologic therapies for RA is small, then there is little value in running a new trial in this area. If EVPI is large enough to suggest a new trial is worthwhile running, then we next need to ask what parameters exactly we would like to collect further information on.

**EV of partial perfect information**

In most models there are many parameters, and we can be more specific about the particular parameters on which it would be useful to have better information. Suppose, for example, that we could split the parameters into two sets, $\theta$ being perhaps the parameters denoting the relative effects of the biologic therapies compared with no treatment on initial ACR, and $\omega$ being the remaining parameters, including, for example, those relating to the rate of decline in HAQ scores in those remaining on treatment, the HAQ to QALY mapping and so on. Thus, we might wish to know: should we conduct further research on the relative efficacies of the biologic therapies, or should we focus more effort on the rest of the natural history model?

The EV of partial perfect information (EVPPi) answers precisely this question [13, 14]. Suppose we are interested in knowing whether it is worthwhile researching the efficacy parameters $\theta$ to eliminate uncertainty in these parameters alone.

For each possible set of parameter values $\theta$ that could occur we find the new best strategy $S^{\text{new}}$ that gives the greatest expected net benefit averaging over the remaining parameters, $\omega$. The calculation then proceeds as for the EVPI calculation. We record the difference in net benefit that arises from using $S^{\text{new}}$ instead of $S^*$. Averaging this loss over all possible sets of $\theta$ that can occur (weighted by how likely they are to occur) gives us the expected loss attaching to uncertainty in parameters $\theta$, termed the EVPPi($\theta$). Note that EVPPi cannot be greater than EVPI, and so if there is no value in eliminating uncertainty in all parameters, there will not be any value in eliminating uncertainty in a subset of them. EVPI is easy to calculate and a first step to identify whether EVPPi calculations are necessary.

The interpretation of EVPPi calculations is subtle as the EVPPi of each parameter, while taking account of uncertainty in the other parameters, still depends on what is assumed about the other parameters [12]. In practice, a hypothetical decision maker wondering whether to fund research on biologics has to consider whether further research on the other parameters $\omega$ might impact on a decision to reduce uncertainty in the efficacy parameters $\theta$. Eventually, the questions can best be seen as being about how much data to collect, rather than on the unachievable ideal of eliminating all uncertainty.

**EV of sample information**

The EV of sample information (EVSI) answers this need, and tells us what is the EV of a decision made after

---

*Fig. 2.* The CEAC for four alternative approaches to increasing uptake of breast cancer screening services. Modified from Ref. [12] with permission from John Wiley & Sons Inc.
research with a given design (e.g. a sample of size \( n \)) has been conducted, generating some hypothetical data \( D \) [15, 16]. For example, this could be a trial of one biologic therapy head-to-head with another with, say, \( n = 100 \) patients on each arm, or it could be a cohort study collecting information on natural history parameters such as HAQ relationships over time after treatment failure. In this final example of an EVI analysis, we imagine a particular study design and sample size for a future study, but we still do not know exactly what results \( D \) we will obtain. We therefore need to average over a distribution of possible data sets \( D \), which would be predicted by our current information on the uncertain parameters \( \theta \) and \( \omega \). For each possible hypothetical data set \( D \) that could arise from a future study of size \( n \) on parameters \( \theta \), we update the probability distributions describing the uncertainty in parameters \( \theta \). These new probability distributions will reflect the reduction in uncertainty resulting from the new data. We then find the new best strategy (biologic therapy) \( S^{new} \) that gives the greatest expected net benefit averaging over the new probability distributions for \( \theta \) and also averaging over the remaining parameters, \( \omega \). The calculation then proceeds as for the EVPI calculation. We record the difference in net benefit that arises from using \( S^{new} \) instead of \( S^* \). Averaging this loss over all possible data sets \( D \) that can occur (weighted by how likely they are to occur) gives us the expected loss attaching to uncertainty in parameters \( \theta \) that can be avoided by carrying out a study of given design (e.g. size \( n \)), termed the EVSI(\( \theta, n \)).

This tells us what the EV is of a decision made after collecting further data from a particular study design (e.g. a head-to-head trial of two biologic therapies with \( n \) patients on each arm). But is it worth the resources required to fund such a study, and would other study designs (e.g. multi-arm trials, trials with different treatment arms, cohort studies, different follow-up times, different sample sizes, etc.) be better value for money?

**Population VOI**

All the above calculations relate to the EV of decisions made regarding a single patient. To use these methods in practice, we must multiply the per patient EV by the number of patients who will enter the particular decision problem. This requires assumptions. The first question is how many new cases enter the decision problem each year. It has been estimated that there are \( \sim 20\,000 \) new cases of RA [17] and \( 2000-5000 \) new cases of PSA [18] per year in the UK. In a specific analysis, of course, the numbers needed would be, for example, the numbers who fail on MTX, or who fail on MTX and at least one biologic.

The more difficult question is the time horizon for the decision. New biologic products for each specific indication are being produced all the time. Technically, the use of the products currently in use and possibly placebo need to be re-evaluated alongside each new product that is produced. NICE currently has a cycle of review that is approximately every 3 years, although some products will be included in a review shortly after being approved through the single technology assessment process. The time period is thus probably between 1 and 3 years. To obtain the population EVI, each of the formulae above needs to be adjusted as follows:

\[
\text{Population EVI} = \text{EVI}^*n^*\text{time horizon}.
\]

Bravo Vergel et al. [4] found in a CEA of etanercept vs infliximab for PsA that the population EVPI was at \( \sim £23\,m \), suggesting that further research to inform this decision could be worthwhile. The partial EVPPIs showed that the primary area of uncertainty where research efforts were of value was the relative short-term effectiveness of the two drugs, in other words a head-to-head trial [4].

**Expected net benefit of sampling**

The final calculation, then, is to evaluate whether it is cost-effective to run a new study (let us say a head-to-head trial of two biologics) and if so to choose the best design for that trial (we will concentrate on sample size per arm, \( n \)). EVSI produces an answer to this second question that is radically different from the standard power calculation. The conventional approach is built on assumptions that are arbitrary, in that specific probabilities of Type 1 and 2 errors are assumed, they ignore cost and population size, and make no recognition of any pre-existing information—such as information already available from indirect comparisons. The Expected Net Benefit of Sampling (ENBS) is simply the difference between the monetarized gain that results from collecting data from the chosen design with sample of size \( n \), and the cost of that study with sample size \( n \):

\[
\text{ENBS} (n) = \text{population EVSI} (n) - \text{cost} (n)
\]

Figure 3 plots expected net benefit of sampling (ENBS) against sample size for the breast cancer screening problem shown in Fig. 2. As samples size increases the EV of a decision made after collecting the sample increases—reaching eventually the EVPI when the sample size is infinite. But costs also rise, and at a rate that settles down to be a fixed cost per patient. ENBS is positive, so there is value in running such a trial. Furthermore, inevitably the ENBS reaches a maximum, which can be considered the optimal sample size. ENBS for a head-to-head trial of two biologics will therefore tell us the optimal sample size for such a trial, and whether it is cost-effective to run a trial at all (if ENBS is positive). ENBS can also be compared for different study designs (e.g. multi-arm trials, trials with different treatment arms, cohort studies, follow-up times) to find the best type of study to fund, including multiple studies (portfolios of research).

ENBS takes into account the existing data on efficacy as this is locked into the distributions of the parameters in the net benefit equation. Clearly, it also takes account of the numbers of patients with the condition. Research on common diseases is therefore prioritized over rare diseases—potentially a controversial point in terms of discrimination. It is worth emphasizing that the costs of the research are not just the crude costs of the trial, but include the additional costs incurred by the health service.
as a result of randomizing some patients to treatments that are suboptimal—particularly to placebo [15].

In fact, these calculations can be used to determine exactly which treatments should be included in the trial, on the basis of existing evidence. EVI calculations may show that randomizing to a placebo arm is not justified by the additional information gained. Figure 4 illustrates a network of evidence on bipolar disorder treatments, involving five active treatments and placebo. Using a simple CEA model, similar to the model used in NICE Technology Assessment 66 [19], we have experimented on various forms of EVSI calculation. Figure 5 shows a plot of the EVSI against the monetarized value of a QALY (willingness-to-pay), when 1500 patients are allocated in equal numbers to each arm. Each line represents a trial that includes different treatment comparisons. The results show that an infinitely sized trial including all seven treatments provides scarcely any more information than a trial of 1500 patients involving the four best treatments: olanzapine, haloperidol, lithium and valproate semisodium. Note that in this context, best means highest in expected net benefit. Placebo does not appear in this list as it is relatively ineffective, quetiapine because it is relatively costly. This approach could be used to determine whether there is value in funding multi-arm trials of biologic therapies, and if so which arms to include.

Some readers may be interested in seeing the formulae for the different forms of EVI calculation. These can be found in the appendix (available as supplementary data at Rheumatology Online), where they are set out in a heavily annotated form.

**CEA models and VOI calculations**

Clearly, VOI calculations begin with a CEA model. It needs to be a CEA model of the specific type—becoming increasing widely used—that fully expresses parameter uncertainty. Models that do not require individual patient simulation are favoured, because the high computational demands of EVPPI and EVSI calculations are then increased exponentially by a third level of simulation, as well as a level of optimization to identify the optimal sample size. While it is true that methods exist that mitigate this problem, they require special skills and software.

Although solutions to computing problems can usually be found, it needs to be emphasized that VOI calculations, like many other techniques, require careful interpretation and sensitivity analysis [12]. It may require a large number of calculations looking at different scenarios, and possible alternative portfolios of research before a trial design can be confidently generated.

The prime requirement, however, is a CEA model that represents a consensus view of the natural history of the disease and the evidence. Clinicians commonly express serious doubts about the assumptions that are made just to achieve a CEA, and are often dismayed that, after
considerable effort, much uncertainty remains. However, VOI analyses are many times more sensitive to assumptions than CEA. They are not only sensitive to changes in the central values, but also extremely sensitive to changes in the degree of uncertainty. It is for this reason that investigators have to be satisfied that the distributions of parameters in the model represent a consensus view; otherwise, they will certainly not be convinced by the research designs that emerge.

**VOI calculations in head-to-head trials of biologics: the policy context**

In the UK context and probably in most jurisdictions, biologics have generally been approved for use in their licensed indications. Their clinical superiority over placebo has been firmly established, though questions remain about how the extent of this superiority relates to disease duration. Since several biologics are in play already, the only effect of collecting head-to-head data will be to knock out one or more accepted treatments from use. (Indeed, in the VOI framework research has no value at all if it cannot change the current optimal decision.)

As noted previously [1], the existing, mostly indirect, evidence does not suggest there are clinically important differences in efficacy, at least between the anti-TNF-α inhibitors. The current evidence probably overstates the uncertainty in this conclusion as it does not incorporate the prior belief we would have had before any evidence was collected, to the effect that if there are differences they are unlikely to be large.

Under these circumstances it is likely that only a large trial would be able to show that one of these treatments is decisively more or less effective than its competitors. There is a distinct possibility, therefore, that at current drug prices, and bearing in mind the ever rising costs of running a trial, a VOI analysis would tell us that head-to-head trials will not be a cost-effective use of public resources. We should also recognize that manufacturers would be unlikely to accept that their product should be removed from the recommended list, unless it was shown that competitors were significantly better at accepted statistical levels, or that there was a decisive difference in cost-effectiveness. It follows that it would be highly unlikely that an inferiority trial would be large enough to affect the decision, unless its purpose was to add a new biologic therapy to the currently accepted set.

Before ending, it is perhaps worth considering the issue of side effects. If there is little to choose between biologics on the basis of efficacy, the decision might be based not only on cost, but also on side-effect profile. This suggests that it may be very important to include a careful analysis of side effects in future models. However, if it is true that the side-effect profile depends on the individual patient, policy makers working at the national level should hesitate before ruling out a treatment on the basis of an average profile, as there may be some patients for whom this is the only drug they can tolerate.

**Rheumatology key messages**

- VOI methods provide guidance on what kind of further research (if any) is cost-effective.
- The use of VOI methods to guide arthritis research requires an economic model supported by consensus.

Supplement: This paper forms part of the supplement ‘Biologic therapies in inflammatory joint diseases: models, evidence and decision making’. This supplement was supported by unrestricted funding from Arthritis Research UK.

Disclosure statement: The authors have declared no conflicts of interest.
Supplementary data

Supplementary data are available at *Rheumatology* Online.

References