A Framework for Cost-Effectiveness Analysis from Clinical Trial Data

Anthony O’Hagan\textsuperscript{1} and John W. Stevens\textsuperscript{2}
\textsuperscript{1}University of Sheffield, UK, and
\textsuperscript{2}Astra Zeneca R&D Charnwood, Loughborough, UK

October 18, 2000

Abstract

We present a general Bayesian framework for cost-effectiveness analysis from clinical trial data. This framework allows for very flexible modelling of both cost and efficacy related trial data. A common cost-effectiveness analysis technique is established for this wide class of models through linking mean efficacy and mean cost to the parameters of any given model.

Examples are given in which efficacy may be measured as a continuous, binary, ordinal or time-to-event outcome, and in which costs are modelled as distributed normally, lognormally, as a mixture or non-parametrically.

A case study is presented, illustrating the methodology and illuminating the role of prior information.

Keywords: Bayesian Analysis; Cost-effectiveness Acceptability Curve; Lognormal Distribution; Net Benefit; Normal Distribution; Outcomes

1 Introduction

The field of Health Economics is growing rapidly, and there is considerable interest from health providers world-wide in assessing the cost-effectiveness of new drugs and treatments. In the UK, the National Institute for Clinical Excellence (NICE) has been established with a specific aim of advising on “current best practice”, and will “address cost as well as clinical effectiveness” of treatments in use by or under consideration for the National Health Service. Several countries have established formal guidelines, including Canada, Australia and the Netherlands. Pharmaceutical companies are beginning to acknowledge that they will need to demonstrate cost-effectiveness in future, rather than simply effectiveness.

A canonical problem in the field is the comparison of two treatments using data from a clinical trial, in which both cost and efficacy are measured
on each patient in each of the two treatment groups. This problem has
been addressed recently by Van Hout et al (1994), Wakker and Klaassen
2000). All of these papers assume normality of the underlying cost and
efficacy data, or at least that the sample size is large enough for sample
means to be normally distributed.

In practice, the data are unlikely to be normally distributed. The clinical
outcome of the trial is often binary, such as eradication or presence of an
ulcer, ordinal such as overall opinion of efficacy on a five-point rating scale,
or a time to event such as death. The distribution of cost data is typically
markedly skewed, and may well be multimodal.

To assume that sample means are normally distributed, relying on the
central limit theorem, is plausible if the sample size is sufficiently large, but
the high degree of skewness in costs means that sample sizes in the hun-
dreds may be needed. Whereas the total sample size in a multi-centre trial
might be this large, in cost-effectiveness analysis it is generally not appro-
priate to treat data from different centres as homogeneous, because of likely
differences in patterns of usage of health care resources. Even if normality
of sample means is reasonable, when the original data are not normal the
sample means will generally not provide the most efficient inferences.

We present a general framework for cost-effectiveness analysis that in-
corporates arbitrary distributions for the data, and so allows the trial data
to be modelled in the most appropriate way.

Our analysis employs the Bayesian approach to statistics. Several au-
thors (for example, Heitjan et al (1999), Jones (1996), Briggs (1999), O’Hagan
et al (2000)) have advocated the use of Bayesian statistics in health eco-
nomics. Its advantages include: offering more intuitive and meaningful in-
ferences; the ability to tackle more complex problems, because of the avail-
ability of powerful computational algorithms; and allowing the use of prior
information in addition to the trial data. In Section 2.4 we discuss the role
of prior information in cost-effectiveness studies.

Much of the literature on comparing the cost-effectiveness of two treat-
ments has been based upon the incremental cost-effectiveness ratio (see,
for example, the theoretical papers of Wakker and Klaassen (1995), Willan
and O’Brien (1996), Laska et al (1997) and practical applications such as
Lacey et al (1999), Pieters et al (1999) and Tennvall et al (1999)). However,
Stinnett and Mullahy (1998) argued for inference and decisions to be based
instead upon a net benefits approach; Van Hout et al (1994) had earlier
proposed the C/E Acceptability Curve (CEAC), which plots the probabil-
ity of a positive net benefit against the threshold cost of a unit increase in
efficacy. O’Hagan et al (2000) endorsed the use of the CEAC and pointed
out that its formulation is intrinsically Bayesian. The first practical uses of
this approach that we are aware of, as the primary analysis protocol for real

Our general framework links the parameters of an arbitrary statistical model for the trial data to the true patient mean efficacies and costs, and hence provides inference about the CEAC and other measures relevant to determining cost-effectiveness. This general framework is presented in Section 2, including illustrations of the generality of the model. A more detailed illustration is presented as a case study in Section 3. Section 4 presents some conclusions.

2 Cost-effectiveness

2.1 Modelling costs and efficacies

We have patient-level data \( D = \{ x_{ij} : i = 1, 2; j = 1, 2, \ldots, n_i \} \) from a clinical trial, where \( x_{ij} \) comprises the observations on subject \( j \) receiving treatment \( i \). In the simplest case, the data are direct measurements of efficacy and cost, whereupon \( x_{ij} \) consists of two numbers, an efficacy measure \( e_{ij} \) and a cost \( c_{ij} \). Then we formally write it as a vector of two elements, \( x_{ij} = (e_{ij}, c_{ij})^T \). We can also allow more general data structures by letting \( x_{ij} \) be an arbitrary vector, perhaps containing more than two numbers. To illustrate why this greater flexibility is desirable, consider a study in which the efficacy is a time to death. These times will naturally be censored because the trial is of only finite length, and patients will be censored at different times because they do not all enter the trial at the same time. In such a case, we will clearly need to record not just the (censored) time to death and the cost but also the censoring time, and therefore \( x_{ij} \) will necessarily contain at least three components.

There are two treatments under consideration, and we are interested in whether treatment 2 is more cost-effective than treatment 1. There are \( n_i \) subjects in treatment group \( i \).

We suppose that a general observation pair \( x \) has a distribution from a family \( \mathcal{F} = \{ f(\cdot | \theta) : \theta \in \Theta \} \) indexed by parameter \( \theta \). The true parameter value for observations under treatment \( i \) is \( \theta_i \). Thus, we assume that the observations \( x_{11}, x_{12}, \ldots, x_{1n_1} \) are a sample from \( f(\cdot | \theta_1) \) and \( x_{21}, x_{22}, \ldots, x_{2n_2} \) are a sample from \( f(\cdot | \theta_2) \). The likelihood is

\[
p(D | \theta_1, \theta_2) = \prod_{i=1}^{2} \prod_{j=1}^{n_i} f(x_{ij} | \theta_i).
\] (1)

In order to assess the cost-effectiveness of treatment 2 relative to treatment 1, we require to compare expected efficacies and expected costs for each treatment. Let the distribution \( f(\cdot | \theta) \) have mean \( \alpha(\theta) = (\mu(\theta), \gamma(\theta))^T \), so that the mean efficacy for treatment \( i \) is \( \mu(\theta_i) \) and the mean cost for treat-
ment $i$ is $\gamma(\theta_i)$. We will use the shorthand notation $\alpha_i = \alpha(\theta_i)$, $\mu_i = \mu(\theta_i)$ and $\gamma_i = \gamma(\theta_i)$.

This is a general framework intended to cover a wide range of specific models. For instance, the efficacy measure might be a continuous quantity such as the level of haemoglobin in patients’ blood, but it is often binary, ordinal or a time to event. Each case will demand a different distributional form $f(\cdot | \theta)$. Similarly, the costs associated with some treatments may reasonably be supposed to be normally distributed, but more often will be skewed, leading again to different distributional forms for the model. Different models will be parameterised differently, and therefore the formulation of $\theta$ will be dependent on the specific application.

Nevertheless, whatever the model and whatever the nature of $\theta$, the efficacy and cost means can be defined as functions $\alpha(\theta)$ of $\theta$, and it is on these that questions of cost-effectiveness depend. To further emphasise the generality of this framework, we present two illustrations in the next two subsections.

2.2 The normal case

The simplest case of the general model is where we have direct observation of efficacy and cost, i.e. $x = (e, c)^T$, and $f(\cdot | \theta)$ is a bivariate normal distribution. Thus, both costs and efficacies are assumed to be normally distributed. The parameter $\theta$ then comprises the mean vector $\alpha = (\mu, \gamma)^T$ and the $2 \times 2$ variance matrix $V$ of efficacy and cost. Formally,

$$f(x | \alpha, \theta) = \frac{1}{2\pi |V|} \exp\left\{-\frac{1}{2}(x - \alpha)^T V^{-1} (x - \alpha)\right\}.$$  \hspace{1cm} (2)

The relationship between the underlying parameters $\theta$ of the model and the parameters $\mu$ and $\gamma$ required for the cost-effectiveness analysis is also particularly simple in this case, since $\mu$ and $\gamma$ are just elements of $\theta$.

This normal framework has been implicitly or explicitly assumed in most of the literature on cost-effectiveness analysis in health economics. See for example Van Hout et al (1994), Wakker and Klaassen (1995), Willan and O’Brien (1996), Laska et al (1997), Heitjan et al (1999). Furthermore, it is common to equate the elements of $\theta$ to the sample variances and covariances, thereby ignoring uncertainty in these parameters: of the references just given, only Laska et al allow in any way for such uncertainty, by making an ad hoc adjustment of normal percentage points to Student $t$ points. A full Bayesian analysis of the normal model is given in O’Hagan et al (1999).

In practice, it is rarely appropriate to assume normality of either costs or efficacies. Cost distributions are almost invariably skew, so that distributions such as the log-normal or gamma would be more appropriate. Efficacy measures in clinical trials are often discrete, the binary case being particu-
larly common. The following scenario illustrates the complexity that could easily arise when we attempt to model the data more faithfully.

### 2.3 A more complex illustration

Now consider a hypothetical trial in which the primary clinical outcome is ordinal, each patient being classified into category 1, 2 or 3 (with 1 being the most successful). We suppose that this outcome is converted into an efficacy measure by assigning value $u_k$ to category $k$, such that $u_1 > u_2 > u_3$. Then the efficacy $e_{ij}$ is $u_k$ if patient $j$ given treatment $i$ is assigned to outcome category $k$. The distribution of $e$ is therefore discrete, with three possible values. If $\pi_k$ is the probability that a patient falls in outcome category $k$ the distribution of $e$ gives probability $\pi_1$ to the value $u_1$, $\pi_2$ to value $u_2$ and $\pi_3$ to value $u_3$. Hence $\mu(\theta) = \pi_1 u_1 + \pi_2 u_2 + \pi_3 u_3$.

Suppose next that in this hypothetical trial we suspect that the cost distribution might be bimodal, with a large proportion of patients having low costs, primarily associated with the treatment and general management, while the remainder will have much larger costs due to the occurrence of additional complications, including adverse events. In order to reflect the way that the distribution of cost for a given patient depends on their outcome category, we let the proportion of patients having low costs in category $k$ be $\phi_k$.

Finally, suppose that the distributions of costs for the low and high cost groups are modelled as gamma distributions with means $\xi_L$ and $\xi_H$ and with variances $\omega_L$ and $\omega_H$. We do not of course observe what ‘cost group’ a patient falls into; the distribution of cost in category $k$ is simply a mixture $\phi_k \text{Ga}(\xi_L, \omega_L) + (1 - \phi_k) \text{Ga}(\xi_H, \omega_H)$. The mean cost becomes $\gamma(\theta) = \sum_{k=1}^{3} \pi_k \{ \phi_k \xi_L + (1 - \phi_k) \xi_H \}$.

The parameters of this model are $\theta = (\pi_1, \pi_2, \phi_1, \phi_2, \phi_3, \xi_L, \xi_H, \omega_L, \omega_H)$. $\pi_3$ does not appear because it is implied by $\pi_3 = 1 - \pi_1 - \pi_2$, but we still have nine parameters in this model. Remember also that this simply describes the general form of the distribution of $e$ and $c$. To define the distributions for both treatments via $\theta_1$ and $\theta_2$ we will have 18 parameters in all.

### 2.4 Prior distributions

In addition to the likelihood function (1), it is necessary in a Bayesian analysis to specify a prior distribution for the parameters. If we represent this as a prior joint density $\pi(\theta_1, \theta_2)$, the posterior joint density is given by Bayes’ theorem as

$$p(\theta_1, \theta_2 | D) \propto f(D | \theta_1, \theta_2) \pi(\theta_1, \theta_2).$$

(3)

As usual, the proportionality symbol expresses the fact that the product of likelihood and prior density on the right hand side of (3) must be scaled
to integrate to one over the range of possible \((\theta_1, \theta_2)\) values. The scaled product is then the posterior joint density \(p(\theta_1, \theta_2 | D)\).

The prior distribution represents information about the distributions of costs and efficacies under the two treatments that is available prior to (or, more generally, in addition to) observing the data \(D\). The ability to incorporate such information is a positive benefit of the Bayesian approach. It allows the analysis to make use of more information, and hence to reach stronger conclusions, than would be available from a frequentist analysis. In clinical studies, it is rarely possible to obtain as much data as one would ideally like. The growth of meta-analysis is evidence of the common wish to bring all the relevant evidence to bear on a question, not just the data from a single trial.

In formulating the prior distribution, a number of questions arise that merit wide discussion among health economists, health care providers and regulatory bodies.

The example analysed in Section 3 can be considered as a contribution to that debate. It has been said that an appropriate prior distribution should incorporate scepticism about the cost-effectiveness of the new treatment. The example shows that the issue of scepticism is more complex than has hitherto been suggested; this is discussed specifically in Section 3.5.

### 2.5 Measures of cost-effectiveness

Treatment 2 is more effective if \(\mu(\theta_2) > \mu(\theta_1)\) and is cheaper if \(\gamma(\theta_2) < \gamma(\theta_1)\). Using the shorthand introduced in Section 2.1, these conditions become \(\mu_2 > \mu_1\) and \(\gamma_2 < \gamma_1\) respectively.

Then the efficacy and cost differentials are defined as \(\Delta_e = \mu_2 - \mu_1\) and \(\Delta_c = \gamma_2 - \gamma_1\). Under the net benefit approach, for a given threshold cost \(K\) per unit efficacy, treatment 2 is more cost-effective than treatment 1 if \(\beta(K) > 0\), where

\[
\beta(K) = K \Delta_e - \Delta_c
\]  

is the net (monetary) benefit of treatment 2 versus treatment 1. The threshold cost \(K\) is the value to the health provider of increasing efficacy for a single patient by one unit.

It is important in health economic evaluations for the measure of efficacy to be such that its value is linear, in the sense that the value of an increase of two units in efficacy is \(2K\). The importance of this requirement is not always appreciated, but the idea of having a simple factor \(K\) to relate cost to efficacy depends upon it. Similarly, the whole idea of a cost-effectiveness ratio makes no sense unless this condition holds. Linearity of the efficacy measure is therefore fundamental to all standard methods of cost-effectiveness analysis.

Thus (4) is the average value, over a large number of patients, of changing from treatment 1 to treatment 2, and hence the assertion that treatment 2
is more cost-effective than treatment 1 if $\beta(K)$ is positive.

Our approach will be Bayesian, and inference about $\beta(K)$ will therefore be based on its posterior distribution given the data $D$. Remembering that $\mu_i = \mu(\theta_i)$ and $\gamma_i = \gamma(\theta_i)$ are functions of the underlying parameters $(\theta_1, \theta_2)$, $\beta(K)$ is also a function of the parameters, and its posterior distribution is therefore derived from their posterior distribution $p(\theta_1, \theta_2 \mid D)$.

The probability of positive net benefit is

$$Q(K) = P(\beta(K) > 0 \mid D),$$

which is evaluated with respect to the posterior distribution. A plot of $Q(K)$ versus $K$ is called the C/E acceptability curve (CEAC). This is a useful guide to decision making in the context where $K$ is not easily determined.

Another guide to decision making is the expected net benefit

$$ENB(K) = E(\beta(K) \mid D) = K E(\Delta_e \mid D) - E(\Delta_c \mid D)$$

and all of these expectations are also evaluated with respect to the posterior distribution.

In the context where a range of values of $K$ is of interest, there will be a corresponding range of values of $ENB(K)$. Then a useful quantity is the break-even cost factor

$$K_0 = E(\Delta_c \mid D) / E(\Delta_e \mid D).$$

The expected net benefit is positive for all $K > K_0$ if $E(\Delta_e \mid D) > 0$, and for all $K < K_0$ if $E(\Delta_e \mid D) < 0$. It is interesting to note that this decision rule is similar to the approach to cost-effectiveness analysis based upon inference about the incremental cost-effectiveness ratio $\rho = \Delta_c / \Delta_e$. Inference about $\rho$ is generally complicated, and for instance its expectation will generally not exist. In contrast, $K_0$ is a ratio of expectations (rather than the expectation of a ratio), always exists (with probability one, for continuous efficacy measures) and is more straightforward to evaluate.

It has been argued by Claxton (1999) that the decision as to whether treatment 2 is more cost-effective than treatment 1 is simply a matter of whether $ENB(K) > 0$. If the expected net benefit is positive then treatment 2 is more cost-effective because to use it in place of treatment 1 leads to an expected cost saving to the health provider to provide the same health benefit. From a purely decision-theoretic viewpoint, if a choice between the two treatments must be taken on the basis of information $D$ then Claxton’s analysis is incontrovertible. The net benefit $\beta(K)$ is effectively a utility function for switching from treatment 1 to treatment 2, relating cost to efficacy through the unit cost $K$. In decision theory, the optimal decision is that which maximises the expected utility.
Nevertheless, other features of the distribution of $\bar{\beta}(K)$ will often be of some relevance. The level of uncertainty about $\beta(K)$, and in particular the probability $1 - Q(K)$ that it is negative, indicate the degree of risk in the decision. When there is an option to postpone a decision on cost-effectiveness in order to gain more data (acknowledging that the extra data, and generally the postponement itself, will be gained at some cost), such considerations will formally enter the decision-making process. The decision will then no longer depend solely on $ENB(K)$ based on data $D$.

Note that we have defined net benefit in the net monetary benefit form, so that the utility is expressed as a cost. It may also be defined as a net health benefit $\Delta_e - \Delta_c/K = \beta(K)/K$, where the utility is expressed as efficacy and the factor $1/K$ is the efficacy that is bought by one unit of money. The formulations are clearly, equivalent and inferences about whether $\bar{\beta}(K) > 0$ are the same as those concerning whether $\beta(K)/K > 0$.

### 2.6 Computation

Even if prior distributions can be assumed to be of standard forms, the joint posterior distribution of $\theta_1$ and $\theta_2$ will in practice be intractable. It will not be possible to derive relevant posterior inferences, such as the quantities $Q(K)$ or $E(\beta(K)|D)$ defined in Section 2.1, analytically. It is therefore necessary to consider computational methods.

Numerical integration (or ‘quadrature’) of the posterior distribution should be feasible in many cases, but becomes impractical when the number of parameters is large. The relatively complex illustration of Section 2.3 has 18 parameters, and this is more or less at the limit of what is possible with numerical integration. If the ordinal outcome had been in more than three categories the number of parameters would definitely rule out quadrature. Even the normal model, which has the smallest practical number of parameters, 10, would entail significant computation if we employed numerical integration.

By far the more favoured approach is Markov chain Monte Carlo (MCMC). O’Hagan et al (1999) found MCMC to be very efficient for the normal model, and it is widely used in other application areas for Bayesian models with hundreds of parameters.

### 3 The TACTIC study

#### 3.1 Data

Our case study is a trial comparing pMDI and Turbuhaler® in the treatment of asthma (Pauwels et al, 1996). This was a multi-centre study, with the largest patient groups being recruited in Canada. We will use the UK data alone for this example. The UK data were analysed using the bivariate
normal model in O’Hagan et al (2000). Here we will employ more realistic modelling to better represent the distribution of cost, and a different efficacy outcome.

The primary motivation for the study was to investigate whether asthmatics, considered adequately treated with CFC inhalers in the form of a pressurised metered dose inhaler (pMDI), could be transferred to Turbuhaler® without decrease in the effect of treatment. We identify pMDI as treatment 1 and Turbuhaler® as treatment 2. As a measure of efficacy we will use the binary outcome of whether a patient experienced no exacerbations during the trial (positive outcome) or experienced at least one exacerbation. The data comprise $x_{ij} = (e_{ij}, c_{ij})^T$, where $e_{ij} = 1$ if patient $j$ in treatment group $i$ experienced no exacerbations, otherwise $e_{ij} = 0$, and where $c_{ij}$ is the cost for this patient.

The data indicate that Turbuhaler® is more effective than pMDI, with 58.1% (36 out of 62) of patients experiencing no exacerbations, while the figure for pMDI is 44.8% (26 out of 58). However, the data do not give a clear indication that one treatment is cheaper than the other. The sample mean and standard deviation of log costs in each treatment and outcome group are as follows.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMDI</td>
<td>5.877 (1.47)</td>
<td>6.021 (1.11)</td>
</tr>
<tr>
<td>Turbuhaler®</td>
<td>6.133 (0.85)</td>
<td>6.370 (0.98)</td>
</tr>
</tbody>
</table>

The mean log costs are lower for pMDI, but pMDI also has larger variances. If we compute a crude sample estimate of mean cost as the exponential of (mean + variance/2), weighted by the sample proportions of positives and negatives, we find 890 for pMDI and 780 for Turbuhaler®.

These estimates are based on assuming log-normal costs. The sample mean costs are 1307 for pMDI and 732 for Turbuhaler®, and these would be the appropriate estimates of population mean cost if we had assumed normally distributed costs. The analysis of these data in O’Hagan et al (1999), based on the simple normal model (2), accordingly finds evidence for Turbuhaler® being cheaper than pMDI. However, examination of the cost data shows clear skewness. A major factor in the large sample mean cost for pMDI is two patients who had very high costs of 3863 and 2915, primarily because both experienced lengthy periods when they were unable to work. Neither patient actually had any exacerbations, which leads to the high variance in the table above for the positive pMDI group, even after the log transformation. We return to this feature of the data when discussing prior information in Section 3.3.
3.2 Model

One key feature of the data is that the efficacy response \( e \) is binary, with 1 denoting the positive outcome of no exacerbations, and 0 being the negative outcome of one or more exacerbations.

It is also important to recognise the skewness in the cost data. It is certainly more appropriate to assume that the cost \( c \) is lognormally distributed than that it is normal. The assumption of log-normality is supported by simple plots of the data. For instance, Figure 1 shows two normal probability plots of the log costs for patients with positive outcomes, in treatment groups 1 and 2, produced using Minitab 12.1 (copyright Minitab Inc., 1998).

![Figure 1. Normal probability plots of log costs — treatment 1, positive outcome (left), and treatment 2, positive outcome (right) — TACTIC data.](image)

The first of these, for positive outcomes in treatment group 1, shows evidence against normality, but this is due primarily to the two outlying costs already referred to. The second plot, for positive outcomes in treatment group 2, is typical of all the other three groups, showing a good fit to the (log-)normal distribution. We concluded that log-normality was an acceptable assumption for the TACTIC cost data.

We now build the joint distribution \( f(., \theta) \) in two stages. First, we suppose that there is a probability \( \phi \) of a positive outcome. Then if \( e = 1 \) we let \( \log c \sim N(\lambda_1, \sigma_1^2) \), while if \( e = 0 \) we let \( \log c \sim N(\lambda_0, \sigma_0^2) \). Thus \( \theta \) comprises the parameters \( \phi, \lambda_0, \lambda_1, \sigma_0^2 \) and \( \sigma_1^2 \), and

\[
f(e, c \mid \phi, \lambda_0, \lambda_1, \sigma_0^2, \sigma_1^2) = \begin{cases} 
\frac{1-\phi}{\sqrt{2\pi \sigma_0^2}} \exp \left\{ -\frac{1}{2\sigma_0^2} (\log c - \lambda_0)^2 \right\} & \text{if } e = 0, \\
\frac{\phi}{\sqrt{2\pi \sigma_1^2}} \exp \left\{ -\frac{1}{2\sigma_1^2} (\log c - \lambda_1)^2 \right\} & \text{if } e = 1.
\end{cases}
\]
Therefore $\mu(\theta) = \phi$, the true proportion of positives, and $\gamma(\theta) = (1 - \phi)\exp(\lambda_0 + \sigma_0^2/2) + \phi\exp(\lambda_1 + \sigma_1^2/2)$, i.e. a weighted average of the means of the lognormal distributions of costs.

3.3 Prior information

The parameters comprise $\theta_1 = (\phi_1, \lambda_{10}, \lambda_{11}, \sigma_{10}^2, \sigma_{11}^2)$ and $\theta_2 = (\phi_2, \lambda_{20}, \lambda_{21}, \sigma_{20}^2, \sigma_{21}^2)$.

The TACTIC trial was started in 1991, and it is not practical now to try to elicit the prior information available before it was conducted. We can, however, look at the reasoning behind the design of the trial, as at least an indication of what prior information then existed.

The trial was originally designed with the specific aim of detecting an increase in the proportion of patients experiencing zero exacerbations from 50% with pMDI to 60% with Turbuhaler®. We wish to formulate a prior distribution for $\phi_1$ and $\phi_2$ to match this prior information, and to express reasonable uncertainty about these proportions. We also wish to incorporate some correlation. If the true proportion of positive outcomes for pMDI were higher than 50%, then we would be inclined to expect the proportion for Turbuhaler® to be higher than 60%. In order to incorporate such a correlation we use log-odds transforms of the $\phi_i$s. Letting $\psi_i = \ln\{(1 - \phi_i)/\phi_i\}$, we propose a bivariate normal prior for $(\psi_1, \psi_2)^T$ with mean $(0, -0.4)$ and variance matrix $\begin{pmatrix} 0.25 & 0.1 \\ 0.1 & 0.25 \end{pmatrix}$. Simulations to check the implications of this prior show that $E(\phi_1) = 0.5$ and $E(\phi_2) = 0.593$, closely agreeing with the specified prior estimates. We also find that the standard deviations of $\phi_1$ and $\phi_2$ are 0.117 and 0.114 respectively, reflecting what amounts in practice to substantial uncertainty about these proportions. However, the correlation results in $P(\phi_2 > \phi_1) = 0.77\bar{7}$, which again is a reasonable prior belief.

The prior joint density of $(\phi_1, \phi_2)$ is shown as a contour plot in Figure 2.
There is no recorded prior information about costs, and it is reasonable to suppose that there would have been little such information at the time of planning the trial. We therefore express weak prior information about the $\lambda_{ik}$ parameters by effectively giving them uniform prior distributions.

We also wish to express weak prior information about the $\sigma^2_{ik}$ parameters. However, the discussion of sample variances in Section 3.1 suggests that this could lead to inappropriate posterior inferences. Although we might have little prior information about costs, we would not expect variances of costs to be very different between the two treatments. Accordingly we formulate a prior distribution that gives weak prior information about variances individually but such that ratios of variances are likely to be reasonably close to unity. Our prior formulation for the $\sigma^2_{ik}$s is hierarchical. Conditional on a hyperparameter $\omega$, we let $\sigma^2_{ik} \sim \omega \chi^2_{20}$. We then give $\omega$ a log-uniform prior distribution. The result is, as required, a weak prior distribution for each $\sigma^2_{ik}$ individually, but the ratio of any such pair of variances has the $F_{20,20}$ distribution. The quartiles of this distribution are 0.74 and 1.36, so that we also have the required effect of the ratios tending to be not too far from unity.

Figure 2. Prior joint density of $\phi_1$ and $\phi_2$, TACTIC data.
3.4 Results

The computations were performed by MCMC using the WinBUGS software package; Spiegelhalter et al (1999).

Figure 3 shows the posterior joint density of $\phi_1$ and $\phi_2$, for comparison with Figure 2. The reduction in uncertainty is shown even more clearly in Figure 4, where the prior and posterior marginal densities are plotted.

![Figure 3. Posterior joint density of $\phi_1$ and $\phi_2$, TACTIC data.](image)

![Figure 4. Prior (dotted lines) and posterior (solid lines) distributions for $\phi_1$ (left) and $\phi_2$ (right), TACTIC data.](image)
The primary quantities for the cost-effectiveness analysis are the differences $\Delta_e$ and $\Delta_c$, whose posterior joint density is shown in Figure 5.

![Figure 5. Posterior distribution on the cost-effectiveness plane, TACTIC data.](image)

This graph has been obtained by bivariate kernel smoothing of 10,000 MCMC sample values, using S-Plus 4.5 (copyright Mathsoft Inc., 1998). It is likely that a larger sample would have produced a plot with smoother contours, but the principal conclusions are clear from Figure 5. Whilst there is evidence that Turbuhaler® is more effective than pMDI, and this is quantified as $P(\Delta_e > 0 \mid D) = 0.96$, the evidence also suggests that Turbuhaler® may be more expensive, with $P(\Delta_c > 0 \mid D) = 0.54$.

Note that O’Hagan et al (1999), who analysed the same cost data with an assumption of normality, found a probability of 0.9 that Turbuhaler® is cheaper than pMDI. The difference is due to two factors. The first is the use of a lognormal distribution for costs here, which gives less weight to the two outlying costs in the pMDI data. The second factor is the prior distribution, and we return to this factor later in this section.

The solid curve in Figure 6 is the CEAC obtained using the specified prior information. From the value of 0.46 at $K = 0$, it rises sharply after $K = 100$ towards the limiting value of 0.96 as $K$ tends to infinity. The probability that Turbuhaler® is more cost-effective exceeds 0.75 for $K$ greater than about 1500, and exceeds 0.9 for $K$ above 5000. Costs are in pounds sterling.
We have $E(\gamma_1 \mid D) = 855.7$, $E(\gamma_2 \mid D) = 864.3$, $E(\phi_1 \mid D) = 0.4576$, $E(\phi_2 \mid D) = 0.5808$. The break-even cost value $K_0$ is therefore $(864.3 - 855.7)/(0.5808 - 0.4576) = 69.8$. Thus, if a decision is required between pMDI and Turbuhaler® , using the decision-theoretic approach of Claxton (1999), Turbuhaler® is preferred for all $K \geq 70$. At this break-even value, the value of $P(\beta(70) > 0 \mid D) = Q(70)$ is around 47.5%. (Asymmetry in the posterior distribution, which would not have been captured by modelling with normal distributions, accounts for the fact that we may find the expected net benefit positive even though the probability that net benefit is positive may be less than 0.5.) Such a decision contrasts with the more conservative approach of demanding substantial evidence, in the form of a $P(\beta(K) > 0 \mid D)$ value of perhaps 0.75, 0.9 or even 0.95 before declaring it adequately demonstrated that treatment 2 is more cost-effective.

Various further analyses were performed, to test the sensitivity of conclusions to the substantive prior information expressed in Section 3.3. It was found that varying the prior information on the mean efficacy parameters $\phi_1$ and $\phi_2$ made negligible difference to conclusions, since the data quite strongly indicate an effectiveness improvement under Turbuhaler®. How-

Figure 6. Cost-effectiveness Acceptability Curves, TACTIC data. Solid line: using specified prior information. Dot-dashed line: using weak prior information.
ever, the effect of the prior information on variances of log costs was much more substantial.

The dot-dashed curve in Figure 6 shows the CEAC under a weak prior specification, where very large prior variances are given to \( \phi_1 \) and \( \phi_2 \) and where the variances of log costs are given independent prior distributions, also with very large variances. The difference from the solid curve is striking for values of \( K \) up to 1000. The CEAC for small \( K \) depends most strongly on the probability that treatment 2 is cheaper than treatment 1. As discussed in Section 3.1, the data alone suggest that Turbuhaler\(^\text{®}\) is cheaper than pMDI, and with weak prior information this feature comes through fully into the posterior distribution. In this alternative analysis \( P(\Delta_c > 0 \mid D) \) is only 0.29, and the CEAC is correspondingly higher for small \( K \). However, we argued in Section 3.1 that this was primarily being driven by two outlying observations, and this was the motivation for the exchangeable prior distribution on the \( \sigma_{ik}^2 \) parameters. The effect of that has been to mitigate the effect of the outliers, so that the remaining evidence in the costs now indicates that Turbuhaler\(^\text{®}\) is slightly more expensive than pMDI.

The difference shows in the different posterior estimates of the \( \sigma_{ik}^2 \) parameters, as shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>( \sigma_{10}^2 )</th>
<th>( \sigma_{11}^2 )</th>
<th>( \sigma_{20}^2 )</th>
<th>( \sigma_{21}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informative prior</td>
<td>1.18</td>
<td>1.71</td>
<td>1.03</td>
<td>0.87</td>
</tr>
<tr>
<td>Weak prior</td>
<td>1.24</td>
<td>2.15</td>
<td>0.96</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Table 1. Estimates of variances of log costs under two prior distributions, TACTIC data.

The effect of the prior information has been to pull these estimates closer together. The largest shift has been in \( \sigma_{11}^2 \), the variance of log costs for positives in the pMDI treatment group, which is where the outlying observations are. The variances have been reduced for both subsets of the pMDI group and increased for both subsets of the Turbuhaler\(^\text{®}\) group. The amount by which the estimates are shifted is directly related to the degrees of freedom, 20 in the informative prior and effectively zero in the weak prior, in the prior distribution for \( \sigma_{ik}^2 \). An intermediate choice, say 5 or 10 degrees of freedom, would imply less strong prior belief in the variances being similar, and hence less shrinkage of the estimates away from their weak prior values towards a common value. A larger degrees of freedom would produce even stronger shrinkage.

The inferences are clearly sensitive to the strength that we specify for this component of the prior information. This is not surprising because in any statistical analysis where the data contain outliers the inferences will tend to be sensitive to how we treat the outliers. In respect of the TACTIC data, we
would argue that it is important to specify realistically strong information so that the estimates are moved appreciably. Whilst one cannot realistically specify a unique value the inferences are reasonably robust to any choice of degrees of freedom between 10 and 50.

3.5 Discussion

An important feature of these data has been the need to model costs carefully. The analysis in O’Hagan et al. (1999) found a probability of 0.9 that Turbuhaler® is cheaper than pMDI, assuming normally distributed costs. This falls to a little over 0.7 using our present analysis with lognormally distributed costs and weak prior information, and to below 0.5 when we add prior information about variances of log-costs. The sensitivity of this probability (and hence of the CEAC generally for low $K$) to the model is marked, and emphasises the importance of careful modelling. It is not enough to employ nonparametric methods or the bootstrap (as in Tambour and Zethraeus (1998)), and so to try to avoid addressing the modelling question. Such an analysis of these data, without prior information to moderate the data, would generally give a relatively high probability that Turbuhaler® is cheaper than pMDI, simply because the sample mean cost is appreciably higher for pMDI. Yet it is our contention that the more correct conclusion for these data is that Turbuhaler® is slightly more likely (probability of 0.54) to be more expensive than pMDI than it is to be cheaper.

The difference between the two analyses presented here also offers an important message with regard to the concept of a sceptical prior distribution. One interpretation of this would be to begin with weak prior information, on the grounds that this favours neither treatment over the other and allows the trial data to determine the posterior distribution. In this example, that approach turns out to be more favourable to Turbuhaler® than our full prior specification, in which we express a prior belief that Turbuhaler® is more effective, but also include a belief that variances of log costs should not vary greatly between treatment and outcome groups. The latter piece of prior information in fact expresses some scepticism about the trial data.

A prior distribution that placed weak prior information on costs (and in particular gave independent weak prior distributions to variances of log costs) but expressed some prior expectation that pMDI would be more effective, would appear to be genuinely sceptical about Turbuhaler®’s cost-effectiveness. But this prior would come to essentially the same conclusion as the weak prior distribution, because of the strength of the data in regard to efficacies. It would still give a more favourable result for Turbuhaler® than our more positive prior distribution.

It is clearly not a simple matter to identify a suitable prior distribution. In this example, prior information that is essentially neutral between the
two treatments is important to include, whereas information which favours one treatment over the other in terms of efficacy is effectively irrelevant.

We should emphasise here that the data analysed in this example are not the full set of data arising from this trial. Analysis of the Canadian data, which comprised the majority of all cases, has already been published and demonstrated clear dominance of Turbuhaler® (Liljas et al., 1997). We analyse the small UK data set here because it exemplifies our methodology and because it shows some features that could be important in other cost-effectiveness studies.

4 Conclusions

We have presented a general Bayesian framework for cost-effectiveness analysis from clinical trial data. The key feature of this framework is a flexible model for the data, in which the mean efficacy μ and mean cost γ associated with a random patient are expressed as functions of the various model parameters.

Decisions about cost-effectiveness depend on the values μ₁, μ₂, γ₁ and γ₂ of these parameters for the two treatments under comparison. In particular, we argue for the value of the expected net benefit \( ENB(K) \) and the probability \( Q(K) \) of cost-effectiveness of treatment 2 for given threshold cost \( K \) associated with unit health benefit. In the context where a unique \( K \) is not available, the break-even unit cost \( K_0 \) and the cost-effectiveness acceptability curve (CEAC) are the comparable decision-support tools.

We have given several examples of how the framework can be applied to a diverse range of clinical trial data. In terms of the efficacy outcome, much of the existing literature on statistical approach to cost-effectiveness assume continuous normally-distributed outcomes, but we show how our framework may be used also for binary, ordinal and time-to-effect outcomes. In terms of costs, we give examples of costs modelled as being distributed normally, lognormally, as a mixture and nonparametrically.

We give a careful and detailed analysis of the TACTIC study, in which we have a binary outcome and lognormal costs. The dataset is small, as is often the case in cost-effectiveness studies, and hence prior information can provide valuable enhancement of the information in the data. We show, in particular, how both the use of non-normal distributions for modelling and the incorporation of prior information on comparability of cost variances are important in moderating the influence of outliers in the data.

We believe firmly that the Bayesian approach offers a variety of significant advantages and has great potential in cost-effectiveness analysis. The issue of the use of (possibly subjective) prior information is a recurring one in applications of Bayesian methodology, and the cost-effectiveness context is no exception. As a contribution to this most important debate, we show
in Section 3.5 that appropriate use of prior information does not necessarily favour new treatments.

Acknowledgements
The authors thank two referees for their stimulating comments.

References


