Int. J. of the Economics of Business,
Vol. 16, No. 3, November 2009, pp. 323–345

Routledge
Taylor & Francis Group

A Bayesian Net Benefit Approach to Cost-effectiveness Analysis in Health Technology Assessment

ELÍAS MORENO, FRANCISCO JAVIER GIRÓN, FRANCISCO JOSÉ VÁZQUEZ-POLO and MIGUEL A. NEGRÍN

ABSTRACT *The economic literature on cost-effectiveness analysis in the context of decisions by health technology assessment agencies assumes as the quantity of interest a linear combination of the mean of the sampling distribution of the effectiveness and the cost. We argue that this is not always reasonable. Our reasons for this assertion are that (i) treatments are compared on the basis of mean values, and for some useful models the mean of the distribution of the cost, which is conditional on the available data, does not exist, and (ii) even for models for which the mean does exist, it might not constitute an accurate reflection of the distribution. This paper presents a general Bayesian cost-effectiveness analysis of a single treatment, where the quantity of interest is the distribution, conditional on the data, of the net benefit. This approach permits a natural extension to several treatments, which enables us to make a statistical comparison. Illustrations with treatment comparisons for real and simulated data are given.*

Key Words: Bayesian Analysis; Cost-effectiveness of a Treatment; logStudent Distribution; Net Benefit; Treatment Comparison.

JEL classifications: I18, C11, C53, H43.

1. Introduction

Statistical techniques in general, and Bayesian tools in particular, have gained popularity in recent years among agencies for health technology assessment who

We thank Prof. Eleanor Morgan (the editor) and an anonymous referee for their helpful comments and suggestions. Financial support by Junta de Andalucía grant SEJ-02814 is gratefully acknowledged. Financial support for this study of the authors EM and FJG was provided in part by Ministerio de Educación y Ciencia (MEC), Spain, grant SEJ2004-02447. Financial support to the authors FJVP and MAN for this study was provided in part by Ministerio de Educación y Ciencia (MEC), Spain, grant SEJ2006-12685. The funding agreement was such as to ensure the authors' independence in writing and publishing the report.

Elías Moreno, Department of Statistics and Operational Research, Faculty of Science, University of Granada, Avenida Fuente Nueva s/n. 18071, Granada, Spain; email: emoreno@ugr.es. Francisco Javier Girón, Department of Statistics and Operational Research, Faculty of Science, Campus de Teatinos, 29071, Malaga, Spain; email: fj_giron@uma.es. Francisco José Vázquez-Polo, Department of Quantitative Methods, Faculty of Economics and Business, Campus de Tafira, 35017, Las Palmas de Gran Canaria, Spain; email: fjvpolo@dmc.ulpgc.es. Miguel A. Negrín, Department of Quantitative Methods, Faculty of Economics and Business, Campus de Tafira, 35017, Las Palmas de Gran Canaria, Spain; email: mnegrin@dmc.ulpgc.es.

1357-1516 Print/1466-1829 Online/09/030323–23

DOI: 10.1080/13571510903227056

need new methodologies for this purpose. The most important agencies for health technology assessment (such as the National Institute for Clinical Excellence (NICE) in the UK, the Food and Drug Administration (FDA) in the US or the Canadian Agency for Drugs and Technologies in Health (CADTH)) recognize the utility of Bayesian statistics in the evaluation of medical devices. In particular, the FDA has recently published the *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials* (FDA, 2006).

It is clear that the problem of comparing several treatments can be regarded as a statistical decision problem[1] (Claxton *et al.*, 2000); in such a context, O'Hagan and Forster (2004: 90) say 'the predictive framework is the most natural way to formulate a real decision problem.' Cooper *et al.* (2002) made the following observation:

> The implementation of decision-theoretic methodology such as the use of the expected value of information could be incorporated into the framework suggested above to identify the cost-utility of conducting future research … the posterior distribution for each parameter obtained from meta-analysis is used to inform the probabilistic decision model. However, in some cases the predictive distribution may be more appropriate; for example, when inferences are made at the individual (unit) level rather than the population (average) level.

Several decision-making measures for cost-effectiveness analysis (CEA) have been proposed, and a brief summary is provided in this section; for a more complete review of the literature, see e.g. Willan and Briggs (2006) and references therein. As usual (Hahn and Whitehead, 2003), let us assume that e_{ij} and c_{ij} are realizations for patient $j \in \{1, \ldots, n_i\}$ on treatment $i \in \{1, \ldots, I\}$, of the random variable effectiveness e and cost c, respectively. Let μ_e and γ_c denote their respective expected values, provided they exist. In most common situations, two competing treatments are analyzed, and hence we will assume that $I = 2$.

The aim of CEA is to relate the clinical effectiveness and health outcomes of services to the net resource costs associated with their use and to identify an efficient use of scarce health care resources. Until just a few years ago, the incremental cost-effectiveness ratio (ICER) was the only quantity of interest in health economics literature. The ICER parameter for comparing treatment i to i' is defined as:

$$ICER = \frac{\gamma_{c_i} - \gamma_{c_{i'}}}{\mu_{e_i} - \mu_{e_{i'}}}. \tag{1}$$

This quantity, however, presents some difficulties for treatment comparison. For instance, the same negative value can indicate preference for the control treatment or for the new treatment, depending on whether the negative sign is due to the incremental effectiveness or the incremental cost. In addition, a ratio statistic poses particular problems for calculating confidence intervals when there is a non-negligible probability of a very small value for the denominator of the ratio.

In response to these difficulties posed by the ICER, a new approach for analyzing the uncertainty in the economic evaluation of health interventions has been proposed, namely the incremental net benefit (INB) parameter (Briggs, 1999; Fenwick *et al.*, 2001; Löthgren and Zethraeus, 2000; Stinnet and Mullahy, 1998). The INB is defined as:

$$INB = R(\mu_{e_i} - \mu_{e_{i'}}) - (\gamma_{c_i} - \gamma_{c_{i'}}). \quad (2)$$

A crucial point in the INB approach is the consideration of the positive deterministic parameter R, which is merely a device for transforming effectiveness into cost, and its meaning is *the value to the health provider of increasing effectiveness for a single patient by one unit* (O'Hagan and Stevens, 2001). In contrast with the ICER, a positive INB value is unambiguously favorable for the new treatment being analyzed.

Since decision-makers encounter difficulties in practice in assessing a unique value for R, interest is focused on the range of values for R for which the Bayesian posterior probability of INB is positive. When this probability is depicted as R varies, the cost-effectiveness acceptability curve (CEAC) is obtained. Such curves have also been suggested as instruments to manage uncertainty concerning the effectiveness and costs of health technologies (Al and van Hout, 2000; Fenwick *et al.*, 2001). However, Koerkamp *et al.* (2007), among others, have recently pointed out significant limitations of CEAC for determining uncertainty in CEA. In the words of the authors, 'these limitations arise because CEAC is unable to distinguish dramatically different joint distributions of incremental cost and effect. … CEACs may mislead policy makers and can incorrectly suggest medical importance.'

Furthermore, we note that the advent of Markov chain Monte Carlo (MCMC) methods has facilitated the development of a relatively large body of recent literature on Bayesian cost-effectiveness analysis (Briggs, 1999; Heitjan *et al.*, 1999; Jones, 1996; O'Hagan and Stevens, 2001, 2002; O'Hagan *et al.*, 2001; Parmagiani, 2002; Spiegelhalter *et al.*, 1994; Stevens *et al.*, 2003).

All the decision-making measures described above are based on the expectation of effectiveness and cost. In this sense, O'Hagan and Stevens (2002) argue that 'from the perspective of the health care provider needing to decide which treatment to apply to the population of patients in their care, it is the mean cost and mean effectiveness over the whole population in which they are interested.' At this point two issues arise: firstly, the mean of a distribution might not be a good summary of the underlying distribution. For example, the mean would be a poor tool as a measure of central tendency for highly asymmetric distributions, and it is known that the distribution of the costs usually presents a high degree of skewness. A related problem is that the mean of a distribution does not necessarily exist. For example, the mean of the cost-effectiveness ratio does not exist when the mean incremental effectiveness is zero. Moreover, as we show in section 4.2, in some cases, the mean of the net benefit may also not exist. This serious drawback can go unnoticed when MCMC computations are carried out and when it is taken for granted that the mean does exist.

'Models in mean' have serious interpretative drawbacks when non-continuous variables are used in the model as, for instance, when effectiveness is a binary indicator (Al and van Hout, 2000; Heitjan and Li, 2004). In fact, when dichotomous effectiveness is considered, the parameter R should be interpreted as the value to the health provider of increasing the probability of success for a single patient by 1%, assuming that this willingness to pay is independent of the initial probability of success (Claxton *et al.*, 2000; Weinstein, 1995).

To overcome the problems with the Bayesian 'models in mean' mentioned above, in this paper we present a Bayesian analysis which offers a general

alternative to the net benefit based on the means. We note that from the distribution, conditional on the observed data, of the variables effectiveness and cost we can draw inferences about the behavior of all the variables involved in the problem. Thus, under this alternative the comparison between treatments is to be based on that between the conditional distributions, given the data, of some functions of the effectiveness and cost of each treatment received by the patients.

We first suggest considering the evaluation of a single treatment through the posterior distribution of a random net benefit variable z. Some characteristics of the conditional distribution of z, which could include the mean, the median, the mode, and the probability that z exceeds a given threshold, are reasonable summaries of this distribution to communicate to the decision-maker. Then, this scheme is extended in order to compare distributions associated with different treatments. This analysis is a generalization of the linear combination of the mean of effectiveness and cost as defined by O'Hagan and Stevens (2001), and is described in section 3.1 endnote.

The rest of the paper is organized as follows. Section 2 presents a general formulation of the cost-effectiveness for a single treatment based on the net benefit approach. In section 3, we consider the problem of treatment comparison. In section 4, we analyze a specific model where the effectiveness is a dichotomous variable, the outcomes of which are success and non-success, with the cost following a lognormal distribution. The reason for doing this is that this model is not only of interest in itself but that it also illustrates how unreliable a cost-effectiveness analysis based on means can be. Section 5 contains some examples with real and simulated data. Finally, section 6 contains a discussion and some concluding remarks.

2. Cost-effectiveness Analysis for a Single Treatment

Following the formulation suggested in Willan *et al.* (2004a, b) and Manca *et al.* (2005), when in a clinical trial cost and effectiveness are available for each patient, the net benefit of patient j is given as:

$$NB_j = R \cdot e_j - c_j.$$

Therefore, a general formulation of the random net benefit for each value of R is the random function z defined as:

$$z = R \cdot e - c \equiv R \times effectiveness - cost. \tag{3}$$

This will be our quantity of interest. By assigning a sampling model for the observable variables effectiveness and cost, which typically depend on certain unknown parameters, and a prior distribution for the parameters, the distribution of the variable z can be derived, although instead of reporting the whole density we would consider summaries of it as these are easier to communicate. Typically, data are available for the effectiveness and cost, and thus we compute the distribution of conditional on the observed data. This posterior distribution forms the basis for our subsequent inferences.

For the sake of clarity, in this paper we consider the case where no deterministic covariates are associated with the patients observed. The important case

where covariates are introduced into the analysis in order to reduce the variability of the effectiveness and cost has already been considered by other authors (Hoch *et al.*, 2002; Vázquez-Polo and Negrín, 2004; Vázquez-Polo *et al.*, 2005a, b; Willan *et al.*, 2004a, b). The Bayesian net benefit approach for this important case will be addressed elsewhere.

2.1 The Distribution of Effectiveness and Cost

Let us suppose that the observed data on the effectiveness and cost of a given treatment option correspond to a group of n patients to which the treatment is being applied, and that a vector with two measurements $\{(e_j, c_j), j = 1, \ldots, n\}$ is provided, where e_j denotes the effectiveness and c_j the cost of the treatment for patient j. Then, consider a sampling model for the variable (e, c), say $f(e, c \mid \xi)$, where ξ represents the unknown unobservable parameters whose interpretation depends on the specific model f. A straightforward way to specify the two-dimensional sampling model $f(e, c \mid \xi)$ is by considering the general decomposition

$$f(e,c \mid \xi) = f(e \mid \theta) f(c \mid e, \psi), \tag{4}$$

which gives the density of e and c as the product of the marginal distribution of e and the distribution of c, conditional on e. The unknown unobservable parameters ξ are separated into θ and ψ indicating which ones affect the distribution of effectiveness and cost, respectively, $\xi = (\theta, \psi)$. The case where cost and effectiveness are independent is a particular case of the above formulation with $f(c \mid e, \psi) = f(c \mid \psi)$.

To complete the specification of the model, we need a prior distribution $\pi(\theta, \psi)$, reflecting our prior belief regarding the behavior of the parameters (θ, ψ). Sometimes the prior information on the parameters is very weak and we are unable to construct a single prior π on this basis. In these circumstances, a reasonable way to proceed is to consider a wide class of priors dictated by our prior belief, and to carry out a Bayesian analysis for the whole class (Berger, 1994; Moreno, 2000). Nevertheless, there is no guarantee that robustness with respect to the priors in the class will be achieved, and unfortunately it must be analyzed case-by-case.

A well-established alternative to Bayesian robust analysis consists of relying on default (or objective) methods (Berger and Bernardo, 1992; Jeffreys, 1961). For many statisticians, default priors are the natural ones to be considered when the prior information is weak, since they are derived solely from the structure of the sampling models. Such automatic priors are used in the following analysis, and also for the illustrations presented in sections 4 and 5, and thus we assume that the prior belief regarding the model parameters is weak. However, the use of prior information is strongly recommended when it is substantial and likely to provide a prior distribution.

Since we have a sample $\{(e_j, c_j), j = 1, \ldots, n\}$ of the variables e and c, we can learn about the parameters θ and ψ through the Bayes theorem. In fact, the posterior distribution of (θ, ψ) can be constructed as

$$\pi(\theta,\psi \mid \text{data}) = \frac{\ell(\theta,\psi)\pi(\theta,\psi)}{\int \ell(\theta,\psi)\pi(\theta,\psi)\, d\theta\, d\psi}, \tag{5}$$

where $\ell(\theta,\psi)=\prod_{j=1}^{n} f\left(e_j \mid \theta\right) f\left(c_j \mid e_j,\psi\right)$ represents the likelihood of (θ, ψ) for the available data, assuming that e_j are conditionally independent given θ, and that c_j are conditionally independent given e_j and ψ. Furthermore, when the parameters θ and ψ are assumed to be independent *a priori*, as in O'Hagan and Stevens (2002), so that $\pi(\theta, \psi) = \pi(\theta)\pi(\psi)$, it follows that they are also independent a posteriori. Indeed, from (4) we have:

$$\pi(\theta,\psi \mid \text{data}) = \frac{\prod_{j=1}^{n} f\left(e_j \mid \theta\right)\pi(\theta)}{\int\prod_{j=1}^{n} f\left(e_j \mid \theta\right)\pi(\theta)\,d\theta} \frac{\prod_{j=1}^{n} f\left(c_j \mid e_j,\psi\right)\pi(\psi)}{\int\prod_{j=1}^{n} f\left(c_j \mid e_j,\psi\right)\pi(\psi)\,d\psi}$$
$$= \pi(\theta \mid \text{data})\pi(\psi \mid \text{data}). \tag{6}$$

The joint distribution of e and c, conditional on the data, is then given by:

$$f(e,c \mid \text{data}) = \int f(e,c \mid \theta,\psi)\pi(\theta,\psi \mid \text{data})\,d\theta\,d\psi \tag{7}$$

which can also be decomposed as:

$$f(e,c \mid \text{data}) = f(e \mid \text{data}) f(c \mid e,\text{data}). \tag{8}$$

2.2 The Distribution of the Net Benefit

From expression (8), it is straightforward to obtain the distribution of the net benefit z, conditional on the data and R. Formally, this is given by:

$$f(z \mid R,\text{data}) = \int f(Re - z \mid e,\text{data}) f(e \mid \text{data})\,de. \tag{9}$$

Some characteristics of this conditional distribution, as a function of R, might be of interest for the decision-making process, for instance:

1. The mean $E(z \mid R,\text{data})$, provided it exists. The mean has been the main quantity of interest in conventional cost-effectiveness analysis. By solving the equation:

   $$E(z \mid R, data) = 0,$$

   for R, we obtain a cutoff value R_0 for which the mean of the net benefit is greater than zero.
2. The median $Me(z \mid R,\text{data})$ provides an interesting robust measure of the net benefit of the treatment. By solving the equation:

   $$Me(z \mid R,\text{data}) = 0$$

   for R, we obtain a cutoff value R_0 for which the median of the net benefit is greater than zero.
3. The probability that z exceeds a given value z_0; this is given by:

   1. $Pr(z \geq z_0 \mid R,\text{data}) = \int f(z \mid R,\text{data})\mathbb{1}_{[z \geq z_0]}(z)\,dz,$

where $\mathbf{1}_A$ represents the indicator function of the set A.
By solving the equation

$$Pr(z \geq z_0 \mid R, \text{data}) = p_{z_0}$$

for R, we obtain the value R_{z_0} for which the net benefit z is greater than the threshold z_0, with probability p_{z_0}. For $z_0 = 0$, the net benefit is positive, with probability p_0. The meaning of R_0 is then the cost per unit of the treatment effectiveness, with a probability p_0.
The representation of the posterior probability

$$Q(R) = Pr(z \geq 0 \mid R, \text{data})$$

as a function of R is termed the *predictive cost-effectiveness acceptability curve* (PCEAC) of the treatment, and this is the posterior probability that the variable z, i.e. the net benefit, is non-negative for the different values of R.
It can be shown that if the posterior distribution of z is a continuous distribution, then the equations $Me(z \mid R,\text{data}) = 0$ and $Pr(z \geq 0 \mid R,\text{data}) = 1/2$ have the same solution for R, and hence the resulting value R_0 is a robust cutoff value.

4. The mode, $Mo(z \mid R,\text{data})$, provided that the density $f(z \mid R,\text{data})$ is unimodal for each R. When e is a discrete variable with a finite support (a very common situation), the density of z is a finite mixture of densities. As this mixture could be multimodal, the mode would then not be very useful.

3. Cost-effectiveness Analysis for Treatment Comparison

In comparing I independent treatments with net benefit variables given by $z_1, \ldots, z_I$, the best net benefit treatment can be found by computing the median curves $Me(z_i \mid R,\text{data})$ and the probability curves $Pr(z_i \geq 0 \mid R,\text{data})$ in the range envisaged for the parameter R and then seeking to identify dominance among the curves or, should the curves cross each other, plotting the pairwise differences between them. Useful complementary information is provided by the median curve $Me(z_i - z_j \mid R,\text{data})$ and the probability curve $Pr(z_i \geq z_j \mid R,\text{data})$ for any pair $i, j = 1, \ldots, 1$.

The effectiveness and the costs of the treatments could be similar; in such a case, this would imply a reduction in the number of parameters and thus a simplification of the cost-effectiveness model. In this section, we explore this issue for two treatments ($I = 2$) by analyzing the case of equal effectiveness.

Suppose we have $\text{data}_1 = \left\{\left(e_{1j_1}, c_{1j_1}\right), j_1 = 1, \ldots, n_1\right\}$ from a group of n_1 patients to whom treatment 1 has been applied, and $\text{data}_2 = \left\{\left(e_{2j_2}, c_{2j_2}\right), j_2 = 1, \ldots, n_2\right\}$ from another group of n_2 patients for treatment 2. Consider a sampling model for the effectiveness and cost of each treatment, which we decompose as before.

$$f(e_i, c_i \mid \xi_i) = f(e_i \mid \theta_i) f(c_i \mid e_i, \psi_i), i = 1, 2.$$

To complete the model, we must consider a prior distribution $\pi_i(\theta_i)\pi_i(\psi_i)$ for the parameters of each model, $i = 1, 2$. Below, we assume that the data on effectiveness $\left\{e_{ij_i}, j_i = 1, \ldots, n_i, i = 1, 2\right\}$ are conditionally independent, in their respective

parameters θ_i, and that $\left\{c_{ij_i}, j_i = 1, \ldots, n_i, i = 1,2\right\}$ are conditionally independent, on e_{ij_i} and ψ_i.

3.1 The Joint Distribution of Effectiveness and Cost

As mentioned above, a reasonable simplification of the joint distribution of effectiveness and cost would be to accept a priori that both treatments are equally effective, that is $\theta_1 = \theta_2 = \theta$. However, since we have data from both treatments this null hypothesis can be put to test in the light of the data. This testing problem can be formulated as a model selection problem between the models

$$M_0: \left\{\prod_{j_2=1}^{n_1} f\left(e_{1j_2} \mid \theta\right) \prod_{j_2=1}^{n_2} f\left(e_{2j_2} \mid \theta\right), \pi(\theta)\right\},$$

$$M_1: \left\{\prod_{j_2=1}^{n_1} f\left(e_{1j_1} \mid \theta_1\right) \prod_{j_2=1}^{n_2} f\left(e_{2j_2} \mid \theta_2\right), \pi(\theta_1)\pi(\theta_2)\right\},$$

where a priori $Pr(M_0) = Pr(M_1) = 1/2$, i.e. the default prior. Standard calculations provide the posterior probability of the models as

$$Pr(M_0 \mid \text{data}) = \frac{B_{01}}{1+B_{01}}, Pr(M_1 \mid \text{data}) = 1 - Pr(M_0 \mid \text{data}),$$

where data = (data$_1$,data$_2$), and B_{01} is the Bayes factor for comparing model M_0 versus M_1,

$$B_{01} = \frac{\int \left\{\prod_{i=1}^{2} \prod_{j=1}^{n_i} f\left(e_{ij} \mid \theta\right)\right\} \pi(\theta) d\theta}{\int \left\{\prod_{j=1}^{n_1} f\left(e_{1j} \mid \theta_1\right)\right\} \pi(\theta_1) d\theta_1 \int \left\{\prod_{j=1}^{n_2} f\left(e_{2j} \mid \theta_2\right)\right\} \pi(\theta_2) d\theta_2}.$$

It should be noted that when the value of $Pr(M_0 \mid \text{data})$ is large, the simplification of the model to be used is effective.

Then, the joint distribution of effectiveness and cost of the treatments, conditional on the whole data set, is a convex combination defined by the posterior probability of the model M_0, and M_1, of the joint distribution of effectiveness and cost, conditional on the data and models, that is

$$f(\mathbf{e}, \mathbf{c} \mid \text{data}) = Pr(M_0 \mid \text{data}) f(\mathbf{e}, \mathbf{c} \mid \text{data}, M_0) + Pr(M_1 \mid \text{data}) f(\mathbf{e}, \mathbf{c} \mid \text{data}, M_1), \tag{10}$$

where $\mathbf{e} = (e_1, e_2)$, $\mathbf{c} = (c_1, c_2)$,

$$f(\mathbf{e}, \mathbf{c} \mid \text{data}, M_0) = f(c_1 \mid e_1, \text{data}_1) f(c_2 \mid e_2, \text{data}_2) f(e_1, e_2 \mid \text{data}), \tag{11}$$

and

$$f(\mathbf{e}, \mathbf{c} \mid \text{data}, M_1) = f(c_1 \mid e_1, \text{data}_1) f(c_2 \mid e_2, \text{data}_2) f(e_1 \mid \text{data}_1) f(e_2 \mid \text{data}_2). \tag{12}$$

Note that in density (11), the effectiveness values are not independent while in density (12) they are. The simple structure of these posterior densities enables straightforward simulations to be performed.

3.2 The Distribution of z_1 and z_2

Assuming that both treatments have a common R to convert effectiveness into cost, the joint density of the net benefit of treatments z_1 and z_2, conditional on the data, is as follows:

$$f(z_1, z_2 \mid R, \text{data}) = Pr(M_0 \mid \text{data}) f(z_1, z_2 \mid R, \text{data}, M_0) + Pr(M_1 \mid \text{data}) f(z_1, z_2 \mid R, \text{data}, M_1), \quad (13)$$

where

$$f(z_1, z_2 \mid R, \text{data}, M_0) = \int f(Re_1 - z_1 \mid e_1, \text{data}_1) f(Re_2 - z_2 \mid e_2, \text{data}_2) \times f(e_1, e_2 \mid \text{data}) de_1 de_2$$

and

$$f(z_1, z_2 \mid R, \text{data}, M_1) = f(z_1 \mid R, \text{data}_1) f(z_2 \mid R, \text{data}_2),$$

where $f(z_i \mid R, \text{data}_i)$ $(i = 1,2)$ is computed using expression (9).

Observe that the posterior independence of (c_1, e_1) and (c_2, e_2), conditional on M_1, implies the posterior independence of z_1 and z_2, conditional on M_1.

Some important summaries for decision making associated with the joint conditional distribution (13) are the following.

i. The probability $Pr(z_1 > z_1 \mid R, \text{data}) = \int f(z_1, z_2 \mid R, \text{data}) 1_{(z_1 > z_2)}(z_1, z_2) dz_1 dz_2$ represents the probability that treatment 1 produces a greater net benefit than does treatment 2, conditional on a value R and the available data. The graphical representation of this probability for the different values of R is called the *predictive cost-effectiveness acceptability curve* (PCEAC) of treatment 1 against treatment 2, and it is a generalization of that described elsewhere as the 'cost-effectiveness acceptability curve' (Fenwick *et al.*, 2001).

The cutoff value of R for which treatment 1 produces a greater net benefit than does treatment 2 with probability γ, for some $0 < \gamma < 1$ might be of interest. This value is given by the solution for R of the equation:

$$Pr(z_1 > z_2 \mid R, \text{data}) = \gamma.$$

As in the usual models proposed for net benefit analysis, two particular situations with respect to R are worthy of attention: $R = 0$, and $R \to +\infty$. It is easily proved that:

$$Pr(z_1 - z_2 > 0 \mid R = 0, \text{data}) = Pr(c_2 - c_1 > 0 \mid \text{data})$$

and

$$\lim_{R\to\infty} Pr(z_1 - z_2 > 0 \mid R, \text{data}) = Pr(e_1 - e_2 > 0 \mid \text{data})$$

ii. The median of the difference in net benefits, $Me(z_1 - z_2 \mid R,\text{data})$, a robust and typically monotonic nonlinear function of R. The value of R for which $Me(z_1 - z_2 \mid R,\text{data}) = 0$ indicates the cutoff point for which treatment 1 is better, in the median, than is treatment 2. This is an important value for decision making.

iii. The expectation of the difference in net benefits $z_1 - z_2$, provided it exists,

$$E(z_1 - z_2 \mid R, \text{data}) = R \cdot E(e_1 - e_2 \mid \text{data}) - E(c_1 - c_2 \mid \text{data}).$$

The value of R for which $E(z_1 - z_2 \mid R,\text{data}) = 0$, is the cutoff value for which treatment 1 is, in mean, better than treatment 2 in terms of cost-effectiveness. This quantity is also known as the break-even cost factor

$$R = \frac{E(c_1 - c_2 \mid \text{data})}{E(e_1 - e_2 \mid \text{data})}.$$

Note that, by construction, the mean of the difference of net benefit is a linear function of R, and thus small variations in the posterior expectation of the difference in effectiveness, say $E(e_1 - e_2 \mid \text{data})$, may result in large variations in the cutoff point of R.[2]

4. A Specific Model: Dichotomous Effectiveness and Lognormal Cost

Suppose that n patients are considered for a clinical trial, and that n_1 of them are randomly assigned to treatment 1, and the rest n_2 to treatment 2. The data consist of a vector of size n_1 where each component gives the result of treatment 1, and a similar vector is applied to treatment 2. For the sake of simplicity, the patients are classified into two categories *success* and *non success* according to whether the treatment has been successful or not, thus measuring the effectiveness of both treatments by a dichotomous random variable (Al and van Hout, 2000; Heitjan and Li, 2004).

The outcomes from the effectiveness of the two treatments are given in Table 1.

The costs associated with these patients are determined, and are shown in Table 2.

Note that some of the entries in Table 1 might be zero, meaning that the corresponding entries in Table 2 would be empty. This does not pose any difficulties for Bayesian analysis of the data.

The statistical model for the data on effectiveness e_i is Bernoulli with an unknown probability of success θ_i, that is $e \mid \theta_i \sim Ber(e_i \mid \theta_i)$. Therefore, the probability mass function of effectiveness of each treatment is

Table 1. Observed effectiveness

	Success	Non Success	Total
Treat. 1	n_{11}	n_{10}	n_1
Treat. 2	n_{21}	n_{20}	n_2

Table 2. Observed costs

	Success	Non Success
Treat. 1	$\left\{c_{11}^{(j)}\right\}$, j = 1, …, n_{11}	$\left\{c_{10}^{(j)}\right\}$, $j = 1, \ldots, n_{10}$
Treat. 2	$\left\{c_{21}^{(j)}\right\}$, $j = 1, \ldots, n_{21}$	$\left\{c_{20}^{(j)}\right\}$, $j = 1, \ldots, n_{20}$

$$f(e_i \mid \theta_i) = \theta_i^{e_i}(1-\theta_i)^{1-e_i}, i = 1,2,$$

where $e_i = 1$ if the patient under treatment i is a *success*, and $e_i = 0$ otherwise.

Further, we assume that the cost of the patients receiving treatment i, conditional on $e_i = 1$, is a random variable following a lognormal distribution, and conditionally on $e_i = 0$ a lognormal distribution with different parameters (O'Hagan and Stevens, 2001; Reed *et al.*, 2003). That is, conditional on the result of treatment e_i, the density function of the cost c_i of treatment i is given by

$$f(c_i \mid e_i, \psi_i) = \begin{cases} \Lambda(c_i \mid \mu_{i1}, \sigma_{i1}^2), if\ e_i = 1, \\ \Lambda(c_i \mid \mu_{i0}, \sigma_{i0}^2), if\ e_i = 0, \end{cases}, i = 1,2,$$

where Λ denotes the lognormal distribution with location and scale parameters $\left(\mu_{i\gamma}, \sigma_{i\gamma}^2\right)$.

4.1 The Distribution of the Net Benefit z of a Single Treatment

Consider an arbitrary treatment and assume, as above, that the effectiveness e follows a Bernoulli model with a probability of success θ, i.e. $e \mid \theta \sim Ber(e \mid \theta)$ and that conditional on $e = 1$ and $e = 0$, the cost follows lognormal distributions, i.e. $f(c \mid e = k) = \Lambda(c \mid \mu_k, \sigma_k^2)$, for $k = 0, 1$.

As above, we assume there is no available prior information on θ, μ_k, σ_k^2 and hence we use default priors, which set independent priors on θ and (μ_k, σ_k^2) as follows. For θ the Jeffreys prior, which is a Beta distribution $Be(1/2, 1/2)$, and for (μ_k, σ_k^2) the reference improper density (Berger and Bernardo, 1992), which is:

$$\pi^N(\mu_k, \sigma_k^2) \propto \frac{1}{\sigma_k^2} \mathbf{1}_{R \times R^+}(\mu_k, \sigma_k^2).$$

If (e_j, c_j), for $j = 1, \ldots, n$, is a sample of size n of effectiveness and costs of the above model and the number of successes is $n_1 = \sum_{j=1}^{n} e_j$, then the posterior distribution of θ, given the data, is:

$$\theta \mid \text{data} \sim Be(\theta \mid n_1 + 1/2, n - n_1 + 1/2). \tag{14}$$

If we now decompose the sample of costs $\mathbf{c} = (c_1, \ldots, c_n)$ into two subsamples c_1 and c_0 corresponding to whether $e = 1$ or $e = 0$ of sizes n_1 and $n_0 = n - n_1$, respectively, and denote by

$$\nu_k = n_k - 1, \bar{v}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} \log c_{kj}, s_k^2 = \frac{1}{\nu_k}\sum_{j=1}^{n_k}\left(\log c_{kj} - \bar{v}_k\right)^2, \text{for } k = 0,1,$$

it then follows that the posterior distribution of μ_k and σ_k^2 is the following normal inverted-gamma distribution

$$\mu_k, \sigma_k^2 \mid \text{data} \sim NIGa\left(\mu_k, \sigma_k^2 \mid \bar{v}_k, \frac{1}{n_k}; \frac{\nu_k}{2}, \frac{\nu_k}{2} s_k^2\right) \tag{15}$$

assuming that $n_k \geq 2$, otherwise, an improper posterior distribution is obtained.

If a conjugate prior is used for μ_k, σ_k^2 instead of the reference prior, the posterior is also a normal inverted-gamma distribution.

To determine the distribution of the net benefit, conditional on the data, we first need the distribution of e and c, conditional on the data. The distribution of the effectiveness e, given the data, follows easily from $e \mid \theta \sim Ber(e \mid \theta)$ and equation (14), and is given by:

$$e \mid \text{data} \sim Ber\left(e \mid \frac{n_1 + 1/2}{n+1}\right).$$

Likewise, the distribution of the cost c, given the data and $e = k$, is

$$f(c \mid e = k, \text{data}) \equiv f(c \mid \bar{v}_k, s_k^2, \nu_k) = \frac{1}{s_k\sqrt{\frac{n_k+1}{n_k}}\sqrt{\nu_k}B(\nu_k/2, 1/2)}$$

$$\times \frac{1}{c}\left(1 + \frac{n_k(\log c - \bar{v}_k)^2}{\nu_k(n_k+1)s_k^2}\right)^{-\left(\frac{\nu+1}{2}\right)}, \text{for } k = 0,1, \tag{16}$$

where $B(v_k/2, 1/2)$ denotes Euler's Beta function of parameters $v_k/2$ and $1/2$. We recognize $f(c \mid \bar{v}_k, s_k^2, \nu_k)$ as the density of a logStudent (hereafter log-t) distribution, and denote it as:

$$f(c \mid \bar{v}_k, s_k^2, \nu_k) \equiv \log\text{-t}\left(c \mid \bar{v}_k, \left(1 + \frac{1}{n_k}\right)s_k^2; \nu_k\right), \text{for } k = 0,1.$$

An important feature of the log-t distribution is that it has very heavy tails, in fact it has no moments (Schmoyeri *et al.*, 1996).

If we write

$$\pi_k = \frac{1/2 + n_k}{1+n},$$

the density of c, conditional on the data, is the following mixture

$$f(c \mid \text{data}) = \sum_{k=0}^{1} \pi_k \log\text{-t}\left(c \mid \bar{v}_k, \left(1 + \frac{1}{n_k}\right)s_k^2; \nu_k\right).$$

From this density, it is easy to obtain the density, conditional on the data, of the net benefit z as:

$$f(z \mid \text{data}) = \pi_0 \log\text{-t}\left(-z \mid \bar{v}_0, \left(1+\frac{1}{n_0}\right)s_0^2; \nu_0\right) + \pi_1 \log\text{-t}\left(R-z \mid \bar{v}_1, \left(1+\frac{1}{n_1}\right)s_1^2; \nu_1\right). \quad (17)$$

Thus the mean of the net benefit z does not exist, but a good summary of (17) is given by its median $Me(z \mid R,\text{data})$ and the posterior probability of having a strictly positive net benefit $Pr(z > 0 \mid R,\text{data})$. Furthermore, note that this posterior probability is bounded by the mean of the effectiveness. In fact:

$$\begin{aligned} Pr(z>0 \mid R,\text{data}) &= Pr(Re-c>0 \mid R,\text{data}) \\ &= Pr(e=1 \mid \text{data})Pr(R>c \mid R,\text{data}) + Pr(e=0 \mid \text{data})Pr(c<0 \mid R,\text{data}) \\ &\leq Pr(e=1 \mid \text{data}) = E(e \mid \text{data}), \end{aligned} \quad (18)$$

which does not depend on R.

The inequality in (18) has two important implications. First, the net benefit of a low-effect treatment, i.e. one with a posterior probability of less than half of being effective, has a probability of less than half of being positive, whatever the value of R. Second, there are values of R for which the posterior probabilities $Pr(z > 0 \mid R,\text{data})$ are as close as desired to $Pr(e = 1 \mid \text{data})$. This implies that the net benefit of an effective treatment (one with a posterior probability of greater than half of being effective) has a probability greater than half of being positive for some value of R.

Another important point is that the interpretation of value R changes when dichotomous effectiveness is considered. In a conventional analysis based on the mean, the parameter R should be interpreted as the value to the health provider of increasing the probability of success for a single patient by 1%, assuming that this willingness to pay is independent of the initial probability of success. However, our model provides a more natural way to understand the parameter R. Following equation (3), the meaning of R is the value to the health provider of obtaining success in the treatment.

The derivation of the distribution of the net benefit for two treatments is provided in the Appendix.

5. Examples

In this section we analyze two examples, the first one based on simulated data, and the second on real data.

5.1 Example 1

This example is based on simulated samples from two populations of patients to whom two treatments are applied. The efficacies are set to $\theta_1 = 0.6$ and $\theta_2 = 0.8$ and the sizes of the simulated samples from the corresponding Bernoulli probability densities are $n_1 = n_2 = 50$. The sample of the patients' costs were simulated

from lognormal distributions with parameters $\mu_1 = 3.3$, $\sigma_1 = 0.5$ (corresponding to a mean cost of 30.72), and $\mu_2 = 3.5$, $\sigma_2 = 0.5$ (mean of 37.52), respectively. Although the model allows for the existence of a correlation between effectiveness and cost, for the purposes of this illustration, independence has been assumed in this simulation. These parameters were chosen in such a way that there is no dominance of one treatment over the other. This generally implies that one treatment is better than the other for an interval of values of R, say $(0,R_0)$, while the opposite is true for the interval (R_0,∞).

For the simulated data, the effectiveness is recorded in Table 3.

The costs associated with these patients were recorded accordingly, and are given in Table 4. Further, the simulated costs of the two treatments were classified according to whether or not the treatment was successful, as described in Table 2. From these data, we estimated the parameters of the four log-t distributions needed to compute the posterior distributions of z_1 and z_2, which are given in the Table 4. This table also contains descriptive summaries, like the mode and the median.

The posterior probabilities of models M_0 and M_1 where $Pr(M_0 \mid \text{data}) = 0.007$ and $Pr(M_1 \mid \text{data}) = 0.993$, thus confirming that the effectiveness values of the two models are significantly different.

Figure 1 illustrates the results of applying the net benefit function to the analysis of the simulated data.

For the net benefit function, an important point, which is related to the analogous behavior described for the case of a single treatment in subsection 4.1, is that the value of R_0 where the *median of differences curve* crosses the zero line indicating that the first treatment is better than the second one for $R < R_0 = 26.9$ (left panel), is exactly the same as the one where the *posterior probabilities curve* of the first treatment being better than the second crosses the 0.5 line; this indicates that both curves basically convey the same information regarding the cost-effectiveness performance of the two treatments.

For the sake of completeness, and to enable comparisons between our net benefit approach to cost-effectiveness and the one based on the weighted differences of the mean parameters of treatments and costs, Figure 2 shows the O'Hagan and Stevens (2001) cost-effectiveness straight line:

Table 3. Simulated effectiveness

	Success	Non Success	Total
Treat. 1	28	22	50
Treat. 2	43	7	50

Table 4. Estimated parameters of the four log-t distributions

	Success	Non Success
Treat. 1	(3.42, 0.40, 27) mode = 26.25 median = 30.62	(3.15,0.35, 21) mode = 20.72 median = 23.34
Treat. 2	(3.55, 0.59, 42) mode = 24.66 median = 34.90	(3.42, 0.33, 6) mode = 27.72 median = 30.47

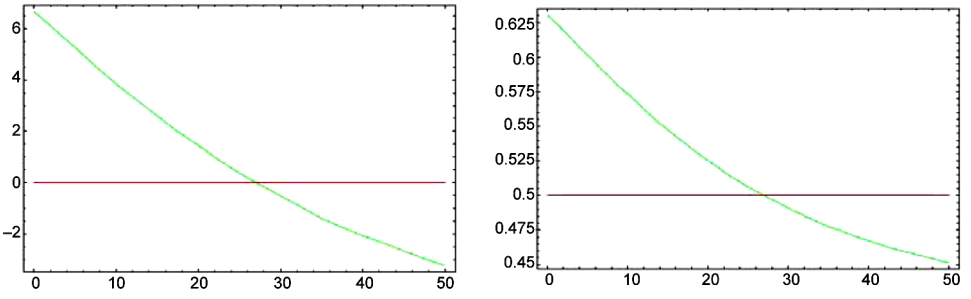

Figure 1. Left panel: plot of the median of the differences between the posterior distributions of z_1 and z_2 against the parameter R. Right panel: PCEAC from data in Example 1.

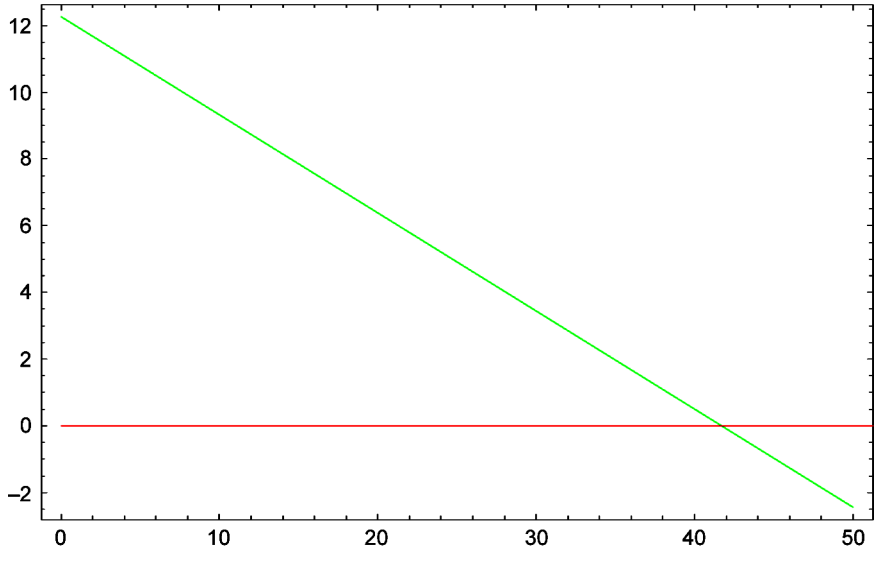

Figure 2. Plot of the linear combination of the differences of posterior means of effectiveness and cost as a function of the parameter R.

$$E(z_1 - z_2 \mid \text{data}) = R\big(E(\theta_1 \mid \text{data}) - E(\theta_2 \mid \text{data})\big) - \big(E(\psi_1 \mid \text{data}) - E(\psi_2 \mid \text{data})\big).$$

In comparing the two approaches, observe that the cutoff point R'_0 of this straight line is markedly greater than the one shown in Figure 1, being in this case $R'_0 = 41.7$. This means that for values of R in the interval (26.9,41.7) the net benefit approach selects treatment 2 as the best, but the mean based posterior analysis still selects treatment 1. Furthermore, the probability of treatment 1 being better than treatment 2 is less than 0.5 in that interval, which we interpret as confirming that our net benefit approach seems to be a more consistent and robust procedure than is the method based on the mean.

5.2 Example 2

The second example is based on real data taken from a clinical trial in which a comparison was made between four highly active antiretroviral treatment

protocols applied to asymptomatic HIV patients (Negrín and Vázquez-Polo, 2006). Although this example includes four treatments, we take two triple combination treatment regimens (d4T+3TC+IND as treatment 1, and AZT+ddl+IND as treatment 2) to present how a cost-effectiveness analysis for two treatments would be carried out.

The effectiveness is measured as the percentage of patients with no detectable virus load. The data on effectiveness, measured by a 0–1 variable, are summarized in Table 5.

Let us first note that the two treatments are very unbalanced, as the sample sizes are 269 and 25, respectively. The posterior probabilities of models M_0 and M_1 are $Pr(M_0 \mid \text{data}) = 0.63$, and $Pr(M_1 \mid \text{data}) = 0.37$, and hence the effectiveness values of both models have a moderate probability of being equal.

There are some large cost values in the first treatment, and a lognormal distribution is required to accommodate them. The estimated parameters of the posterior distribution of the costs of the two treatments are given in Table 6.

A simple intuitive analysis of the above data may run as follows. The estimated parameters of the cost distribution of treatment 1 are slightly greater than those of treatment 2. The objective Bayes estimate of the effectiveness of the first treatment, say $\hat{\theta}_1 = 0.65$, is slightly greater than that for the second treatment $\hat{\theta}_2 = 0.60$, and so it seems that the first treatment has slightly better effectiveness than the second one, but that it also presents higher costs. Thus, it is not clear which of the two treatments is preferable.

However, the net benefit analysis shows that the second treatment outperforms the first one for all values of the parameter R. In fact, the median of the distribution, conditional on the data, of the difference between the net benefits of treatments $z_1 - z_2$, and the posterior probability of the second treatment being better than the first one, convey the same message, namely that the second treatment is uniformly better than the first one for all values of the parameter R. Moreover, the posterior probability of the first treatment being better than the second one is quite small and it varies within a very narrow interval (0.246, 0.229), as R varies. These assertions are illustrated in Figure 3.

Table 5. Observed effectiveness

	Success	Non Success	Total
Treat. 1	175	94	269
Treat. 2	15	10	25

Table 6. Estimated parameters of the four log-t distributions

	Success	Non Success
Treat. 1	(8.86, 0.14, 174) mode = 6950.59 median = 7078.46	(8.85, 0.17, 93) mode = 6780.98 median = 6982.41
Treat. 2	(8.71, 0.08, 14) mode = 6017.26 median = 6051.28	(8.77, 0.09, 9) mode = 6395.34 median = 6446.83

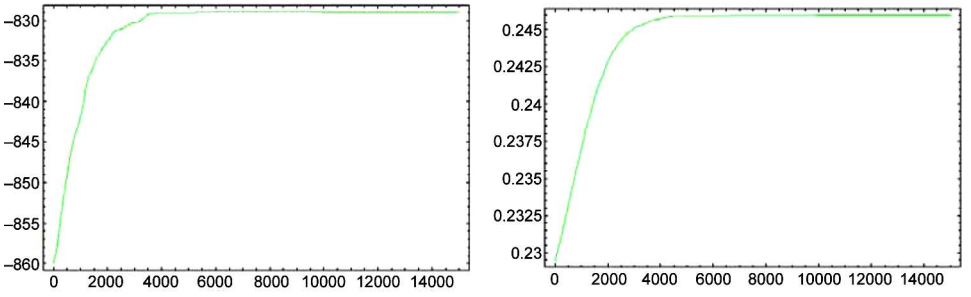

Figure 3. Left panel: plot of the median of the differences between the posterior distributions of z_1 and z_2 against the parameter R. Right panel: PCEAC from data in Example 2.

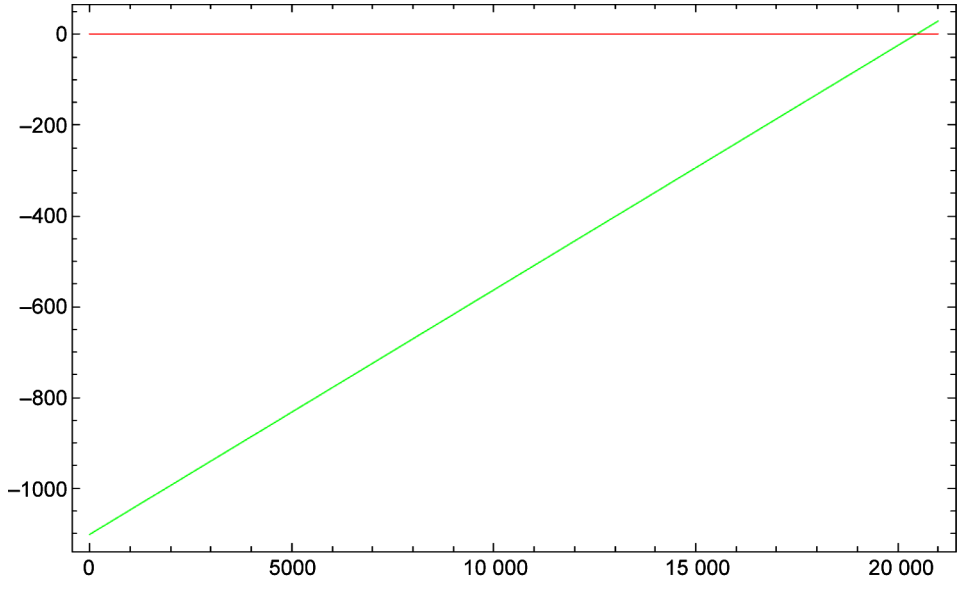

Figure 4. Plot of the linear combination of the differences of posterior means of effectiveness and cost as a function of the parameter R.

It is interesting to see that the above conclusions do not hold when the standard analysis based on the mean is performed. Figure 4 illustrates the results of using this analysis of these real data. Note that the mean-based approach produces the same answer as the net benefit approach for those values of R in the interval (0, 20457), but the opposite is true for those values of R in the interval (20457, ∞).

The disagreement between the two methods is highlighted in the effectiveness analysis for sufficiently large values of R. The standard mean-based analysis considers as the measure of effectiveness the quantity $Pr(\theta_1 > \theta_2 \mid \text{data})$, and the general net benefit the quantity $Pr(e_1 > e_2 \mid \text{data})$. From expression (13) we can compute that $Pr(\theta_1 > \theta_2 \mid \text{data}) = 0.70$. However, our method states that $Pr(e_1 > e_2 \mid \text{data}) = 0.24$, which follows from:

$$Pr(e_1 > e_2 \mid \text{data}) = Pr(e_1 = 1, e_2 = 0 \mid \text{data}) =$$
$$= Pr(M_0 \mid \text{data}) \cdot f(e_1 = 1, e_2 = 0 \mid \text{data}) + Pr(M_1 \mid \text{data}) \cdot f(e_1 = 1 \mid \text{data}) \cdot$$
$$f(e_2 = 0 \mid \text{data}) = 0.24$$

This somewhat unrealistic situation, where $R \to \infty$, illustrates an important point, which is whether the interest for the decision maker should focus on $Pr(e_1 > e_2 \mid \text{data})$, or on the variation in the probabilities of success $Pr(\theta_1 > \theta_2 \mid \text{data})$, which may be of little consequence in the outcome. We believe it is the outcome variable $(e_1 > e_2 \mid \text{data})$ that matters.

6. Discussion

The novel point addressed in this paper is that the cost-effectiveness of a treatment ought to be measured in terms of the distribution of a net benefit function that depends on a crucial parameter R, and the distribution of the variables cost and effectiveness. The parameter R is a tool for expressing the effectiveness of the treatment in units of cost, and the sampling model assumed for effectiveness and costs determines the model for the net benefit, conditional on each value of R.

The net benefit function typically used in the literature on the subject only depends on the mean of the sampling distribution of the effectiveness and cost. This is not necessarily appropriate, especially when the underlying distributions are asymmetric. Furthermore, the incremental net benefit is by construction a linear function of R, regardless of the model considered for effectiveness and cost, which does not seem to be reasonable.

However, by computing the whole distribution of the net benefit, as has been done in the present study, we can choose the characteristics of the distribution considered to be appropriate summaries for it. This choice appears to be highly dependent on the specific problem being addressed. Of course, the use of robust measures of location or concentration of the distribution (such as the median or the probability of particular, relevant events) are recommended, in particular when dealing with long tailed distributions for costs. Certainly, these measures are not necessarily linear functions of R, as expected, but the form of the function depends on the model considered. The examples considered suggest that the analysis of cost-effectiveness based on the distribution, conditional on the data, of the net benefit function is superior to that based on the posterior means of the sampling distributions of effectiveness and cost.

We emphasize that the method proposed here is capable not only of measuring the net benefit of a treatment, but also of comparing treatments, by doing what is natural: comparing their net benefit distributions. It is interesting to observe that, in our examples, the median of the distribution of the difference between the net benefit of the treatments and the posterior probability that one treatment is better than the other convey the same message, so they are consistent.

We have considered a specific model where the effectiveness is measured through a dichotomous variable. This model has consequences in the interpretation of the parameter R, where its meaning is the willingness to pay for obtaining success in the treatment. This meaning is more natural than the meaning of the analysis based on the mean, i.e. the willingness to pay for increasing the effectiveness by 1%. Moreover, the comparison of the effectiveness of two treatments is based on natural units of effectiveness. In our analysis, the interest is not in

comparing the probability of success of treatments, but in how the difference in success probabilities of the treatments influences the outcome for the patients. This fact can have dramatic consequences on decision making, as shown in Example 2.

In this study, the net benefit associated with a treatment is defined by a linear combination of effectiveness and cost. Although this is reasonable, other combinations are certainly possible and worth exploring. Another open question is the analysis of the net benefit of a treatment (i) when the effectiveness is measured by a more general variable than the 0–1 considered here; and (ii) when covariates are included in the sampling model for the effectiveness and cost in order to reduce the variability of the variables.

Notes

1. Even though Bayesian decision analytical modeling methods in health economics is a relatively new research area, there is a large body of literature in other fields based on utility decision theory (Cooper *et al.*, 2003; Spiegelhalter *et al.*, 1994) or expected value of information (Claxton and Lu, 2004; Eckerman and Willan, 2007).
2. Let us cite the analysis developed by O'Hagan and Stevens (2001), as a particular case of ours, when $E(e_i \mid \theta_i) = \theta_i$ and $E(c_i \mid \psi_i) = \psi_i$ and the posterior expectations of e and c both exist. From Fubini's theorem we have

$$E(e_i \mid \text{data}) = \int e_i f(e_i \mid \text{data}) de_i = \iint e_i f(e_i \mid \theta_i)\pi(\theta_i \mid \text{data}) de_i d\theta_i =$$
$$= \int E(e_i \mid \theta_i)\pi(\theta_i \mid \text{data}) d\theta_i = E(\theta_i \mid \text{data}),$$

and, in a similar way, $E(c_i \mid \text{data}) = E(\psi_i \mid \text{data})$, so that the expectation $E(z_1 - z_2 \mid R, \text{data})$ is found to be

$$E(z_1 - z_2 \mid R, \text{data}) = R\Delta e - \Delta c,$$

where $\Delta e = E(\theta_1 \mid \text{data}) - E(\theta_2 \mid \text{data})$ and $\Delta c = E(\psi_1 \mid \text{data}) - E(\psi_2 \mid \text{data})$, which is exactly the formula (4) used by O'Hagan and Stevens (2001).

References

Al, Maiwenn J. and van Hout, Ben A. (2000) A Bayesian approach to economic analyses of clinical trials: the case of stenting versus balloon angioplasty, *Health Economics,* 9(7), pp. 599–609.

Berger, James O. (1994) An overview on robust Bayesian analysis (with discussion), *Test,* 3(1), pp. 5–125.

Berger, James O. and Bernardo, J.M. (1992) On the development of reference priors, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds), *Bayesian Statistics* 4 (Oxford: Oxford University Press).

Briggs, Andrew A. (1999) A Bayesian approach to stochastic cost-effectiveness analysis, *Health Economics,* 8, pp. 257–61.

Claxton, Karl, Ades, A.E., and Lu, G. (2004) Expected value of sample information in medical decision making, *Medical Decision Making,* 24(2), pp. 207–27.

Claxton, Karl, Lacey, Larry F., and Walker, Stephen G. (2000) Selecting treatments: a decision theoretic approach, *Journal of the Royal Statistical Society: Series A,* 163(2), pp. 211–25.

Cooper, Nicola J., Sutton, Alex J., and Abrams, Keith R. (2002) Decision analytical economic modelling within a Bayesian framework: application to prophylactic antibiotics use for caesarean section, *Statistical Methods in Medical Research,* 11, pp. 491–512.

Cooper, Nicola J., Sutton, Alex J., Mugford, Miranda, and Abrams, Keith R. (2003) Use of Bayesian Markov Chain Montecarlo methods to model cost-of-illness data, *Medical Decision Making,* 23(1), pp. 38–53.

Eckermann, Simon and Willan, Andrew R. (2007) Expected value of information and decision making in HTA, *Health Economics,* 16(2), pp. 195–209.

FDA (2006) Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials, *Draft Guidance for Industry and FDA Staff.*

Fenwick, Elisabeth, Claxton, Karl, and Sculpher, Mark (2001) Representing uncertainty: the role of cost-effectiveness acceptability curves, *Health Economics,* 10(8), pp. 779–87.

Hahn, Seokyung and Whitehead, Anne (2003) An illustration of the modelling of costs and efficacy data from a clinical trial, *Statistics in Medicine,* 22(6), pp. 1009–24.

Heitjan, Daniel F. and Li, Huiling (2004) Bayesian estimation of cost-effectiveness: an importance-sampling approach, *Health Economics,* 13(2), pp. 191–98.

Heitjan, Daniel F., Moskowitz, Alan J., and Whang, William (1999) Bayesian estimation of cost-effectiveness ratios from clinical trials, *Health Economics,* 8(3), pp. 191–201.

Hoch, Jeffery, Briggs, Andrew, and Willan, Andrew R. (2002) Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis, *Health Economics,* 11(5), pp. 415–30.

Jeffreys, Harold (1961) *Theory of Probability* (Oxford: Oxford University Press)

Jones, David A. (1996) Bayesian approach to the economic evaluation of health care technologies, in: Spiker, B. (Ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials* (Philadelphia: Lippincott-Raven).

Koerkamp, Bas G., Hunink, Myriam M.G., Stijnen, Theo, Hammitt, James K., Kuntz, Karen M., and Weinstein, Milton C. (2007) Limitations of acceptability curves for presenting uncertainty in cost-effectiveness analysis, *Medical Decision Making,* 27(2), pp. 101–11.

Löthgren, Mickael and Zethraeus, Niklas (2000) Definition, interpretation and calculation of cost-effectiveness acceptability curves, *Health Economics,* 9(7), pp. 623–30.

Manca, Andrea, Rice, Nigel, Sculpher, Mark J., and Briggs, Andrew H. (2005) Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models, *Health Economics,* 14(5), pp. 471–85.

Moreno, Elias (2000) Global Bayesian robustness, in: D. Ríos and F. Ruggeri (Eds), *Bayesian Robustness* (New York: Springer).

Negrín, Miguel A. and Vázquez-Polo, Francisco-José (2006) Bayesian cost-effectiveness analysis with two measures of effectiveness: the cost-effectiveness acceptability plane, *Health Economics,* 15(4), pp. 363–72.

O'Hagan, Anthony and Forster, Jon (2004) *Kendall's Advanced Theory of Statistics. Vol. 2B. Bayesian Inference* (London: Edward Arnold).

O'Hagan, Anthony and Stevens, John W. (2001) A framework for cost-effectiveness analysis from clinical trial data, *Health Economics,* 10(4), pp. 302–15.

O'Hagan, Anthony and Stevens, John W. (2002) Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies, *Statistical Methods in Medical Medical Research,* 11(6), pp. 469–90.

O'Hagan, Anthony, Stevens, John W., and Montmartin, Jacques (2001) Bayesian cost-effectiveness analysis from clinical trial data, *Statistics in Medicine,* 20(5), pp. 733–53.

Parmigiani, Giovanni (2002) *Modelling in Medical Decision Making: A Bayesian Approach (Statistics in Practice)* (Chichester: Wiley).

Reed, Shelby S., Dillingham, Peter W., Briggs, Andrew H., Veenstra, David L., and Sullivan, Sean D. (2003) A Bayesian approach to aid in formulary decision making: incorporating institution-specific cost-effectiveness data with clinical trial results, *Medical Decision Making,* 23, pp. 252–64.

Schmoyeri, R.L., Beauchamp, J.J., Brandt, C.C., and Hoffman, F.O. (1996) Difficulties with the lognormal model in mean estimation and testing, *Environmental and Ecological Statistics,* 3, pp. 81–97.

Spiegelhalter, D.J., Feedman, L.S., and Parmar, M.K.B. (1994) Bayesian approaches to randomized trials (with discussion), *Journal of the Royal Statistical Society. Series A,* 157(3), pp. 357–456.

Stevens, John W., O'Hagan, Anthony, and Miller, Paul (2003) Case study in the Bayesian analysis of a cost-effectiveness trial in the evaluation of health care technologies: depression, *Pharmaceutical Statistics,* 2(1), pp. 51–68.

Stinnett, Aaron A. and Mullahy, John (1998) Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis, *Medical Decision Making,* 18(2), pp. S68–80.

Vázquez-Polo, Francisco-José and Negrín, Miguel A. (2004) Incorporating patients' characteristics in cost-effectiveness studies with clinical trial data. A flexible Bayesian approach, *Statistics and Operations Research Transactions,* 28(1), pp. 87–108.

Vázquez-Polo, Francisco-José, Negrín, Miguel A., and González, Beatriz (2005a) Using covariates to reduce uncertainty in the economic evaluation of clinical trial data, *Health Economics,* 14(6), pp. 545–57.

Vázquez-Polo, Francisco-José, Negrín, Miguel A., Badía, Xavier, and Roset, Montse (2005b) Bayesian regression models for cost-effectiveness analysis, *European Journal of Health Economics*, 50, pp. 45–52.
Weinstein, Milton C. (1995) From cost-effectiveness ratios to resource allocation: where to draw the line?, in: F.A. Sloan (Ed.), *Valuing Health Care* (Cambridge: Cambridge University Press).
Willan, Andrew R. and Briggs, Andrew H. (2006) *Statistical Analysis of Cost-effectiveness Data* (Chichester: Wiley).
Willan, Andrew, Briggs, Andrew, and Hoch, Jeffery (2004a) Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data, *Health Economics,* 13(5), pp. 461–75.
Willan, Andrew, Lin, D.Y., and Manca, Andrea (2004b) Regression methods for cost-effectiveness analysis with censored-data, *Statistics in Medicine,* 24(1), pp. 131–45.

Appendix: The Net Benefit Distribution for Two Treatments

Using the preceding results, we can now provide explicit expressions for the distributions involved in formulas (10)–(12) for the data summarized in Tables 1 and 2. The conditional distributions of e_i given the data$_i$, for $i = 1, 2$, are Bernoulli

$$e_i \mid \text{data}_i \sim Ber\left(e_i \mid \frac{n_{i1} + 1/2}{n_i + 1}\right).$$

By now writing

$$\nu_{ik} = n_{ik} - 1, \bar{\upsilon}_{ik} = \frac{1}{n_{ik}} \sum_{j=1}^{n_{ik}} \log c_{ik}^{(j)}, s_{ik}^2 = \frac{1}{\nu_{ik}} \sum_{j=1}^{n_{ik}} \left(\log c_{ik}^{(j)} - \bar{\upsilon}_{ik}\right)^2, i = 1,2, k = 1,0,$$

the posterior densities of c_i given e_i can be written as:

$$f(c_i \mid e_i = k, \text{data}_i) = \log\text{-}t\left(c_i \mid \bar{\upsilon}_{ik}, \left(1 + \frac{1}{n_{ik}}\right) s_{ik}^2; \nu_{ik}\right), i = 1,2.$$

To obtain the full conditional distribution in formula (10) it is necessary to compute the posterior probability of models M_0 and M_1, and the joint predictive distribution of the pair (e_1, e_2) conditional on M_0 and the whole data.

Application to the Bernoulli case of the model selection procedure described in section 3 renders

$$Pr(M_0 \mid \text{data}) = \frac{B_{01}(\text{data})}{1 + \text{B}_{01}(\text{data})'}$$

where B_{01} (data) is the Bayes factor of model M_0 versus M_1, which is expressed as

$$B_{01}(\text{data}) = \frac{B(n_{\cdot 1} + 1/2, n_{\cdot 0} + 1/2)}{B(n_{11} + 1/2, n_{10} + 1/2)B(n_{21} + 1/2, n_{20} + 1/2)}$$

where $n_1 = n_{11} + n_{21}$, $n_0 = n_{10} + n_{20}$, and B(·,·) denotes Euler's Beta function.

Finally, in order to compute the posterior distribution of the pair (e_1, e_2) under model M_0, we first need the posterior distribution of the common parameter θ, which is given as:

$$\theta \mid \text{data}, M_0 \sim Be(\theta \mid n_{\cdot 1} + 1/2, n_{\cdot 0} + 1/2).$$

As $f(e_1, e_2 \mid \theta) = \theta^{e_1 + e_2}(1-\theta)^{2 - e_1 + e_2}$, integrating out θ yields the desired distribution of the pair (e_1, e_2), conditional on the data, as

$$f(e_1, e_2 \mid \text{data}) = \frac{B(n_{\cdot 1} + e_1 + e_2 + 1/2, n_{\cdot 0} - e_1 - e_2 + 5/2)}{B(n_{\cdot 1} + 1/2, n_{\cdot 0} + 1/2)}$$

where $e_1 = 0,1$, and $e_2 = 0,1$.

If we define the following probabilities for $i = 1,2$, and $k = 0,1$,

$$\begin{aligned}\pi_{ik} &= Pr(M_0 \mid \text{data}) f(e_1 = 2 - i, e_2 = k \mid \text{data}) \\ &\quad + Pr(M_1 \mid \text{data}) f(e_1 = 2 - i \mid \text{data}_1) f(e_2 = k \mid \text{data}_2)\end{aligned}$$

then the conditional density of (c_1, c_2), given the data, is the mixture

$$\begin{aligned}f(c_1, c_2 \mid \text{data}) = \sum_{i=1}^{2} \sum_{k=0}^{1} \pi_{ij} \, &\text{log-t}\left(c_1 \mid \bar{v}_{1k}, \left(1 + \frac{1}{n_{1k}}\right) s_{1k}^2; \nu_{1k}\right) \\ &\times \text{log-t}\left(c_2 \mid \bar{v}_{2k}, \left(1 + \frac{1}{n_{2k}}\right) s_{2k}^2; \nu_{2k}\right).\end{aligned}$$

The net benefit of the treatments (z_1, z_2) is now

$$z_i = Re_i - c_i = \begin{cases} R - c_i, \text{if } e_i = 1 \\ c_i, \text{if } e_i = 0 \end{cases}, i = 1,2.$$

By adapting the results in subsection 3.2 to this dichotomous case, it follows that the density of the pair (z_1, z_2), conditional on R and the data, is given by the mixture:

$$\begin{aligned}f(z_1, z_2 \mid R, \text{data}) = \sum_{i=1}^{2} \sum_{k=0}^{1} \pi_{ij} \, &\text{log-t}\left(R(2-i) - z_1 \mid \bar{v}_{1k}, \left(1 + \frac{1}{n_{1k}}\right) s_{1k}^2; \nu_{1k}\right) \\ &\times \text{log-t}\left(Rk - z_2 \mid \bar{v}_{2k}, \left(1 + \frac{1}{n_{2k}}\right) s_{2k}^2; \nu_{2k}\right).\end{aligned}$$

From this joint distribution the marginals can be obtained as

$$f(z_i \mid R, \text{data}) = \pi_i \log\text{-t}\left(R - z_i \mid \bar{v}_{i1}, \left(1 + \frac{1}{n_{i1}}\right) s_{i1}^2; \nu_{i1}\right)$$
$$\times (1 - \pi_i) \log\text{-t}\left(-z_i \mid \bar{v}_{i0}, \left(1 + \frac{1}{n_{i0}}\right) s_{i0}^2; \nu_{i0}\right).$$

where $\pi_1 = \pi_{11} + \pi_{10}$ and $\pi_2 = \pi_{21} + \pi_{20}$.

Although the expectations $E(z_1 \mid R, \text{data})$ and $E(z_2 \mid R, \text{data})$ do not exist, the medians as a function of R can be easily computed, and represent sensible summaries of the corresponding marginal densities, as shown in section 5. We can also compute the median of the difference between the net benefits, that is $Me(z_1 - z_2 \mid R, \text{data})$, which provides useful information about the range of values of R for which one of the treatments, in terms of the median, is more cost-effective than the other.

Furthermore, the probability $Pr(z_1 > z_2 \mid R, \text{data})$, which is generally a monotonous function of R, might tell us the values of R for which this probability exceeds a given threshold, for instance 1/2.