SORT 33 (2) July-December 2009, 193-212

Eliciting expert opinion for cost-effectiveness analysis: a flexible family of prior distributions

M. Martel, M. A. Negrín and F. J. Vázquez-Polo*

University of Las Palmas de Gran Canaria

Abstract

The Bayesian approach to statistics has been growing rapidly in popularity as an alternative to the classical approach in the economic evaluation of health technologies, due to the significant benefits it affords. One of the most important advantages of Bayesian methods is their incorporation of prior information. Thus, use is made of a greater amount of information, and so stronger results are obtained than with frequentist methods. However, since Stevens and O'Hagan (2002) showed that the elicitation of a prior distribution on the parameters of interest plays a crucial role in a Bayesian cost-effectiveness analysis, relatively few papers have addressed this issue.
In a cost-effectiveness analysis, the parameters of interest are the mean efficacy and mean cost of each treatment. The most common prior structure for these two parameters is the bivariate normal structure. In this paper, we study the use of a more general (and flexible) family of prior distributions for the parameters. In particular, we assume that the conditional densities of the parameters are all normal.
The model is validated using data of a real clinical trial. The posterior distributions have been simulated using Markov Chain Monte Carlo techniques.

MSC: 97K80, 62F15.

Keywords: Bayesian analysis, cost-effectiveness, prior information, elicitation, conditionally specified distributions.

1 Introduction

Spiegelhalter, Feedman and Parmar (1994) argued the use of Bayesian methodology as a formal basis for applying external evidence in cost-effectiveness analysis (CEA). Since then, many authors have discussed the advantages of this methodology versus the classical or frequentist approach (Briggs, 1999; Heitjan, Moskowitz and William, 1999; Fry-

Address for correspondence: Department of Quantitative Methods, Faculty of Economics, University of Las Palmas de Gran Canaria, E-35017 Las Palmas de Gran Canaria. Canary Islands, Spain. {mmartel or mnegrin or fjvpolo}@dmc.ulpgc.es

Received: April 2009

Accepted: September 2009

back, Chinnis and Ulvila, 2001; O'Hagan, Stevens and Montmartin, 2001; Vázquez-Polo and Negrín, 2004; among others).

The incorporation of prior information allows Bayesian methods to access more information and so to produce stronger inferences. Stevens and O'Hagan (2002) discuss the advantages of incorporating prior information in cost-effectiveness analysis of clinical trial data, exploring mechanisms to safeguard scientific rigour in the use of prior information. Since it has become available, a number of different techniques have been developed to elicit or extract prior information from experts (O'Hagan, Buck, Daneshkhah, Eiser, Garthwaite, Jenkinson, Oakley and Rakow, 2006). However, few studies have addressed these techniques in the area of health economics (Fenwick, Palmer, Claxton, Sculpher, Abrams and Sutton, 2006; Smith and Marshall, 2006; Leal, Wordsworth, Legood and Blair, 2007). In the present paper we aim to collaborate in promoting the use of Bayesian analysis by proposing a general and flexible method to incorporate prior information by means of conditionally specified distributions.

CEA is a form of economic evaluation that examines both the costs and health outcomes of alternative health technologies or treatments. The most prevalent measures for the comparison of treatments are the incremental cost-effectiveness ratio (ICER), the incremental net benefit (INB) and the cost-effectiveness acceptability curve (CEAC).

The ICER is defined by:

$$ICER = \frac{\gamma_1 - \gamma_0}{\mu_1 - \mu_0} = \frac{\Delta\gamma}{\Delta\mu}, \tag{1}$$

where γ_j and μ_j are the average cost and effectiveness under treatment j (1, new; and 0, for the current or control treatment), respectively.

The INB of treatment 1 versus treatment 0 is defined as

$$INB(R_c) = R_c \cdot \Delta\mu - \Delta\gamma, \tag{2}$$

for each R_c, which is interpreted by O'Hagan and Stevens (2001) as the cost that decision-makers are willing to accept in order to increase the effectiveness of the treatment applied by one unit. Thus, analyzing whether the alternative treatment is more cost effective than the control treatment is equivalent to determining whether $INB(R_c)$ is positive.

In practice, it is not a simple matter for the decision-maker to determine a single R_c, and so a CEAC is constructed. This curve provides a graphical representation of the probability of the alternative treatment being preferred ($\Pr(INB(R_c) > 0)$) for each value R_c.

We focus on the normal case. Classical cost-effectiveness analysis and most published Bayesian studies assume normality in the distributions of cost and effectiveness (Willan and O'Brien, 1996; Laska, Meisner and Siegel, 1997; Stinnett and Mullahy, 1998; Tambour, Zethraeus and Johannesson, 1998; Heitjan *et al.*, 1999; Briggs, 1999).

Although efficacy outcome data can be binary and patient cost data are likely to be right skewed, the central limit theorem guarantees for sufficiently large sample sizes that the means will be normally distributed. Löthgren and Zethraeus (2000) affirm that "the normal distribution result is valid whether or not the individual cost and effect distributions are normal. The more skewed and non-normal the individual distribution is, the larger sample sizes are needed for the normal distribution approximation to be valid".

A Bayesian analysis of the normal case was examined by O'Hagan *et al.* (2001), who considered the patient level data $\{x_{ij} : i = 1,2,\ldots,n_j; j = 0,1\}$ from a clinical trial, where $x_{ij} = (e_{ij}, c_{ij})$ consists of an effectiveness measures e_{ij} and an associated cost c_{ij}. The index j is used to denote the treatment and n_j denotes the sample size for each treatment j.

We denote by $f(x_{ij}|\mu_j, \gamma_j, \Sigma_j)$ the parametric distribution generating data x_{ij} from treatment j. The parameters of this function are the mean cost (γ_j), the mean efficacy (μ_j) and the variance-covariance matrix Σ_j. Then the likelihood is:

$$\ell(\bar{x}|\mu_0, \gamma_0, \Sigma_0; \mu_1, \gamma_1, \Sigma_1) = \prod_{j=0}^{1} \prod_{i=1}^{n_j} f(x_{ij}|\mu_j, \gamma_j, \Sigma_j). \tag{3}$$

It is assumed that $f(x_{ij}|\mu_j, \gamma_j, \Sigma_j)$ is a bivariate normal distribution for each treatment j

$$f(x_{ij}|\alpha_j, \Sigma_j) = (2\pi|\Sigma_j|)^{-1/2} \exp\left\{-\frac{1}{2}(x_{ij} - \alpha_j)^{\mathsf{T}} \Sigma_j^{-1} (x_{ij} - \alpha_j)\right\}, \tag{4}$$

where $\alpha_j = (\mu_j, \gamma_j)$, i.e. the mean effectiveness and cost for treatment j, respectively.

A Bayesian analysis of model (4) requires the specification of a prior distribution on the parameters. Quantifying the expert's opinion as a probability distribution is a difficult task, and the method presented is intended to help the expert perform the task in a way that is as easy, rigorous and computationally allowable as possible.

A convenient class of prior distributions is a general conditional-conjugate prior. Specifically, O'Hagan *et al.* (2001) assume a bivariate normal distribution for α_j and an inverse Wishart prior distribution for the variance matrices Σ_j.

Although the bivariate normal prior distribution is general and convenient it does present some limitations. For example, the correlation between variables is independent of the values of the variables and it is a unimodal distribution. In this paper, we study the use of a more general family of prior distributions for the parameters of interest. In particular, we assume that the conditional density of μ_j for a given γ_j and the conditional density of γ_j for a given μ_j are both normal. This assumption is manifestly different from one of classical bivariate normality with its familiar elliptical contours. The utility of conditionally specified priors has been explored in other areas, such as the analysis of insurance claims (Sarabia and Gómez-Déniz, 2008; Sarabia, Castillo, Gómez-Déniz and Vázquez-Polo, 2005).

The paper is organized as follows: Section 2 presents the normal case of cost-effectiveness analysis with prior distributions based on a conditional specification. In Section 3 some examples are given to show that the methodology is readily applicable. We use a practical application with real data from a clinical trial, comparing two alternative treatments for asymptomatic HIV patients. Markov Chain Monte Carlo (MCMC) procedures are used to simulate the posterior distribution. Section 4 presents a discussion of the results obtained and some conclusions are then drawn.

2 Bayesian cost-effectiveness analysis with prior distributions based on conditional specification

Our basic prior formulation for model (4) assumes that the joint distribution factorizes as

$$\pi(\alpha_0, \alpha_1, \Sigma_0, \Sigma_1) = \pi(\alpha_0) \cdot \pi(\alpha_1) \cdot \pi(\Sigma_0) \cdot \pi(\Sigma_1). \tag{5}$$

That is, we assume independence between treatments and between the means (α_j) and the variance matrices (Σ_j). Inverse Wishart prior distributions are assumed for the variance matrices Σ_0 and Σ_1. Specifically, we take $\Sigma_j \sim IW(A_j, f_j)$ the prior density of which is

$$\pi(\Sigma_j) \propto |\Sigma_j|^{(-f_j+3)/2} \exp\left\{tr(\Sigma_j^{-1} A_j)/2\right\},$$

over the space of positive-definite 2×2 matrices. Thus f_j is the prior degrees of freedom parameter and the prior expectation of Σ_j is $(f_j - 3)^{-1} A_j$, provided $f_j > 3$.

It is reasonable to assume a prior normal distribution of μ_j for a given γ_j and of γ_j for a given μ_j. A bivariate normal distribution was proposed by O'Hagan *et al.* (2001), but that is only a particular case with normal conditionals.

Castillo and Galambos (1989) showed the specification of the class of all bivariate densities with normal conditionals. We seek to obtain all joint densities $\pi(\mu, \gamma)$ such that every conditional density of μ given γ is normal with mean $\delta_1(\gamma)$ and variance $\sigma_1^2(\gamma)$ (which may depend on γ) and every conditional density of γ given μ with mean $\delta_2(\mu)$ and variance $\sigma_2^2(\mu)$ (which may depend on μ).

The above authors found that all the bivariate densities with normal conditionals are those of the form

$$\pi(\mu, \gamma) = \exp\left\{ \left(1, \mu, \mu^2\right) \begin{pmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1 \\ \gamma \\ \gamma^2 \end{pmatrix} \right\}. \tag{6}$$

The conditional expectations and variances are:

$$\begin{aligned}
\mathrm{E}[\mu|\gamma] &= -\frac{m_{10}+m_{11}\cdot\gamma+m_{12}\cdot\gamma^2}{2(m_{20}+m_{21}\cdot\gamma+m_{22}\cdot\gamma^2)},\\
\mathrm{Var}[\mu|\gamma] &= -\frac{1}{2(m_{20}+m_{21}\cdot\gamma+m_{22}\cdot\gamma^2)},\\
\mathrm{E}[\gamma|\mu] &= -\frac{m_{01}+m_{11}\cdot\mu+m_{21}\cdot\mu^2}{2(m_{02}+m_{12}\cdot\mu+m_{22}\cdot\mu^2)},\\
\mathrm{Var}[\gamma|\mu] &= -\frac{1}{2(m_{02}+m_{12}\cdot\mu+m_{22}\cdot\mu^2)}.
\end{aligned} \tag{7}$$

The distribution with density of the form (6) is an eight-parameter family of densities. The coefficient m_{00} is a normalizing constant that is determined by the other coefficients m and the requirement that the density should integrate to 1. Additionally, some restrictions on the coefficients m should be considered to ensure a positive value for the variances. This point is well illustrated by Arnold, Castillo and Sarabia (2001b).

This family of prior densities is very flexible and contains, as particular cases, many other distributions similar to that proposed in Bayesian literature for cost-effectiveness analysis. Thus, this family represents a significant extension to the usual priors considered. Its interest is twofold. Firstly, due to its conditioned-conjugancy property, it is very easy to simulate MCMC samples from posterior densities using Gibbs sampling. The practical application of the procedure presented in this paper is in accordance with Winkler (2001), as regards ease of use and ready acceptance, bearing in mind that the factors of expertise and prior knowledge can be incorporated into the computations (Malakoff, 1999). Secondly, this class of prior distributions contains a huge catalogue of highlighted prior densities (Spiegelhalter*et al.*, 1994 and Spiegelhalter, Myles, Jones and Abrams, 2000a). For instance, if we are willing to accept improper priors, then conditions for the above parameters (m_{00} among others) are not required. Sceptical priors about treatment effects are also easily elicited by making $E[\mu|\gamma]$ equal to zero and allowing a high degree of spread using the expression of the variances.

Thus, we encounter a great variety of distributions for different values of the m parameters. Some of these distributions are markedly different from classical bivariate normal densities. We now show the values of the m parameters for some particular cases.

- **Independence:** If we assume prior independence between the mean of the effectiveness (μ) and the mean of the costs (γ) for a given treatment, the conditional distributions do not depend on the other parameter, and the conditional expectations and variances will be of the form:

$$\begin{aligned}
\mathrm{E}[\mu|\gamma] &= \mathrm{E}[\mu] = -\frac{m_{10}}{2\cdot m_{20}},\\
\mathrm{Var}[\mu|\gamma] &= \mathrm{Var}[\mu] = -\frac{1}{2\cdot m_{20}},\\
\mathrm{E}[\gamma|\mu] &= \mathrm{E}[\gamma] = -\frac{m_{01}}{2\cdot m_{02}},\\
\mathrm{Var}[\gamma|\mu] &= \mathrm{Var}[\gamma] = -\frac{1}{2\cdot m_{02}}.
\end{aligned} \tag{8}$$

Thus, the conditions for independence are that the m's satisfy the following conditions:

$$m_{11} = m_{12} = m_{21} = m_{22} = 0, \quad m_{20} < 0, \quad m_{02} < 0. \tag{9}$$

- **Bivariate normal distribution**

  Another important case of bivariate distribution with normal conditional is that of the bivariate normal distribution. The bivariate normal prior distribution in Bayesian CEA was proposed by O'Hagan *et al.* (2001) and it is included as a particular case of the conditionally specified prior.
  For the terms μ and γ this can be expressed as

$$\pi(\mu,\gamma|\delta_\mu,\delta_\gamma,\sigma_\mu,\sigma_\gamma,\rho) = \frac{1}{2\pi\sigma_\mu\sigma_\gamma\sqrt{1-\rho^2}}\exp\left\{\frac{Q}{2(1-\rho^2)}\right\},$$

  where σ_μ and σ_γ are the expectations of the mean effectiveness and mean cost respectively, σ_μ and σ_γ are the standard deviation, ρ is the Spearman rho correlation coefficient, and Q is the quadratic expression

$$Q = \frac{(\mu-\delta_\mu)^2}{\sigma_\mu^2} - \frac{2\rho(\mu-\delta_\mu)(\gamma-\delta_\gamma)}{\sigma_\mu\sigma_\gamma} + \frac{(\gamma-\delta_\gamma)^2}{\sigma_\gamma^2}.$$

  The conditional distributions are normal with mean and variance

$$\begin{aligned}
\mathrm{E}(\mu|\gamma) &= \delta_\mu + \frac{\rho\sigma_\mu}{\sigma_\gamma}(\gamma-\delta_\gamma),\\
\mathrm{Var}(\mu|\gamma) &= \sigma_\mu^2(1-\rho^2),\\
\mathrm{E}(\gamma|\mu) &= \delta_\gamma + \frac{\rho\sigma_\gamma}{\sigma_\mu}(\mu-\delta_\mu),\\
\mathrm{Var}(\gamma|\mu) &= \sigma_\gamma^2(1-\rho^2).
\end{aligned} \tag{10}$$

The prior information can be elicited from expressions (7) and (10). Thus, the condition for the bivariate normal distribution is that the m's satisfy the following conditions (Arnold *et al.*, 2001a,b).

$$m_{12} = m_{21} = m_{22} = 0, m_{20} < 0, m_{02} < 0 \quad \text{and} \quad m_{11}^2 < 4m_{02}m_{20}. \tag{11}$$

Of course the use of conditional normal distributions is not the only way to elicit a bivariate normal distribution. In this sense Sarmanov (1966) and Ting Lee (1996) propose a family of bivariate distributions that can be elicited taking into account the marginal distributions.

- **A more general case:**

The improvement obtained from the use of conditionally specified priors is the wide range of prior information that may be elicited. For example, there are some combinations of m's that have non-normal marginal densities. In particular, bimodal or even trimodal densities may be encountered. These distributions must satisfy the conditions for integrability of (6) (Gelman and Meng, 1991, Arnold *et al.*, 2000, Arnold, Castillo and Sarabia, 2001a).

$$m_{22} < 0, \quad 4m_{22}m_{02} > m_{12}^2, \quad 4m_{22}m_{20} > m_{21}^2. \tag{12}$$

However, there is a price to be paid for the flexibility of our prior structure, namely that there are eight hyperparameters to assess. Given the difficulties of eliciting a high-dimensional joint probability distribution, we concentrate on eliciting some important summaries of the distribution, such as means and variances. We recommend the method for matching conditional moments proposed by Arnold, Castillo and Sarabia (1998). For a conditionally specified prior such as (6-7), we can try to match conditional moments, whose approximate values will be supplied by expert opinion. In our analysis, at least eight conditional moments are needed to determine all the hyperparameters. However, this might not be enough to determine the prior information and so it is preferable for the expert to supply more than eight conditional moments. We recognize that it is unlikely that such prior values will be consistent and what we propose is to select a prior of the form (6) that will have conditional moments that are minimally disparate from those provided a priori by the experts.

Let us assume that prior assessed values for the conditional means and variances of the effectiveness and cost are obtained for several different given values of the cost and effectiveness, respectively.

$$\begin{aligned} \mathrm{E}[\mu|\gamma_{p_1}] &= e_{p_1} \quad \forall p_1 = 1,2,\ldots,P_1. \\ \mathrm{Var}[\mu|\gamma_{p_2}] &= var(e)_{p_2} \quad \forall p_2 = 1,2,\ldots,P_2. \\ \mathrm{E}[\gamma|\mu_{p_3}] &= c_{p_3} \quad \forall p_3 = 1,2,\ldots,P_3. \\ \mathrm{Var}[\gamma|\mu_{p_4}] &= var(c)_{p_4} \quad \forall p_4 = 1,2,\ldots,P_4. \end{aligned} \tag{13}$$

where $P_1 + P_2 + P_3 + P_4 \geq 8$.

A unique solution for this system of equations is unlikely to be possible for any choice of the eight hyperparameters. A possible solution is to allow any deviance between the prior conditional moment and the knowledge of the expert. We define as the objective function the sum of the squared deviances (Arnold, Castillo and Sarabia, 1999). The hyperparameters are obtained by minimizing the objective function subject to constraints (12). A LINGO® code containing the procedure used in this article is available from the authors upon request. The prior distribution obtained must be checked by the experts so as not to obtain local minima in the optimization.

The choice of subjective priors is thus a difficult one, and requires the expert to take into account both psychological and behavioural aspects in order to obtain a coherent prior distribution (Baranski and Petrusic, 1994; Yates, 1990; Yaniv, Yates and Smith, 1991; among others). On the one hand, psychological studies have shown how well subjects make estimates and how different techniques of elicitation may produce different responses (Winkler, 1967; Staël von Holstein, 1970. An excellent review of this question was performed by Hogarth, 1975). Furthermore, many pioneering empirical studies (Kahneman and Tversky, 1972; Chesley, 1978; among others) have shown that training and maturity help an expert quantify prior probabilities.

Systematic methods of elicitation are presented in Kadane, Dickey, Winkler, Smith and Peters (1980), Garthwaite, Kadane and O'Hagan (2005) and a recent review of the question was provided by O'Hagan *et al.* (2006). We present a plausible alternative procedure from which it may be realistic to expect the elicitation of the (conditioned) prior mean and variance or other quantities; the specification based conditioned distribution theory may then be used to obtain a full specification of the prior distribution. Inspired by Berger (1994), we propose to use a class of plausible priors to ensure that as many reasonable priors as possible are included. Such a class does not require a strong mathematical training to be elicited and the priors are computationally manageable.

One practical situation where this more general prior distribution can be useful is that of the bimodal case. A bimodal distribution (or in general multimodal distribution) typically indicates that the distribution is in fact the sum of two or more different distributions, each with a single notable peak. Suppose that the treatment involves some risk, and there is a probability that complications

may appear. In that case, the effectiveness could be lower and the costs higher (McIntosh, Ramsey, Berry and Urban, 2001; Viviane and Barkun, 2008). If it is possible to distinguish and to record which patients suffer complications during the study, it would be plausible to propose as the likelihood of the data a mixture of bivariate normal distributions where the weight of each distribution is the probability of complications (Negrín and Vázquez-Polo, 2006). However it is not often easy to define a complication. Two possible solutions would be either to fix an arbitrary threshold cost (or effectiveness) to define the complication, or to approximate the probability of a complication using finite-mixture distributions (Diebolt and Robert, 1994). Conditionally specified distribution can be useful when the presence of complications is not clearly delimited. In this case the prior information can be modelled by a bimodal bivariate distribution, using conditionally specified prior distributions.

3 An example with real data

The data used in this section were obtained from a real clinical trial developed in 1999 in which a comparison was made between various highly active antiretroviral treatment protocols applied to asymptomatic HIV patients (Pinto, López, Badía, Corna and Benavides, 2000).

We only considered the direct costs (of drugs, medical visits and diagnostic tests), and as the effectiveness we considered the improvement in the quality of life, measured using the visual analogue scale (VAS) of the EQ-5D instrument (Brooks, 1996). In particular, we used the variation in the VAS by the end of the study. Cost and effectiveness values were recorded six months after the beginning of the study.

In this exercise, two three-way treatment protocols were compared. The first of these (d4T + 3TC + IND) combined the drugs estavudine (d4T), lamivudine (3TC) and indinavir (IND); the second treatment protocol (d4T + ddl + IND) combined estavudine (d4T), didanosine (ddl) and indinavir (IND).

Table 1 summarizes the statistical data. The d4T + ddl + IND treatment was more costly than the d4T + 3TC + IND treatment, by an average of 164.82 euros. The d4T + ddl + IND treatment was on average more effective, with an improvement in the patients' quality of life of 4.94 units, while those who were given the d4T + 3TC + IND treatment only experienced a VAS improvement of 4.56 units.

Table 1: *Statistical summary of costs (in thousands of euros) and effectiveness (change in VAS).*

	d4T + 3TC + IND		d4T + ddl + IND	
Statistical measure	Cost	Change in VAS	Cost	Change in VAS
Mean	7.142	4.56	7.307	4.94
s.d.	0.001573	15.17	0.001720	13.98
n	$n_0 = 268$		$n_1 = 93$	

For a fully Bayesian analysis, priors for the parameters of interest must be specified. Prior information was obtained from three experts who participated in the study and reflects the reasoning behind the design of the trial. The elicitation method was implemented in an interactive computer program. The computer displays assessment questions and the expert types in answers that reflect his opinion. At any point in the elicitation process, the expert can review the coherence of his probability judgments. Prior distributions derived from experts' consensus are displayed graphically to be reviewed.

Our Bayesian experiment requires the elicitation of normal distributions. A univariate normal distribution is characterized by two parameters, the mean and the variance. The mean (which coincides with the median) and the first and third quartile were requested of the experts in an elicitation process to obtain the prior mean and variance of the parameters of interest. Kadane and Wolfson (1998) suggest that the expert is only comfortable providing the mean and quartiles. Normal distributions were fitted to similar fitting procedures using percentile judgements in Cooke and Slijkhuis (2003), Denham and Mengersen (2007) and Kennedy, Anderson, O'Hagan, Lomas, Woodward, Gosling and Heinemeyer (2008).

- Independence

  The first analysis shows the independence case. For the purpose of this analysis, we took the design of the study to imply prior expectations for the parameters of interest. The experts' expectations show an average of 4.5 units of effectiveness for the control treatment (d4T + 3TC + IND), with a prior variance of 2.25. For the same treatment, the design anticipates an average cost of 5000 euros, with a variance of 4. The value of the m parameters is calculated directly, in the knowledge of the prior mean and variance of effectiveness and cost. For this prior information, the values are:

$$m_{01} = 1.25, m_{02} = -0.125, m_{10} = 2, m_{11} = 0,$$
$$m_{12} = 0, m_{20} = -0.2222, m_{21} = 0, m_{22} = 0.$$

  The elicitation process is very similar for the new treatment (d4T + ddl + IND). In this case, the experts considered this treatment to be less effective, with an average of 4 units of effectiveness and a prior variance of 2.5. They also expected it to be more expensive, with a prior mean cost of 6000 euros, and a variance of 6.25. The values of the m parameters for this treatment are

$$m_{01} = 0.96, m_{02} = -0.08, m_{10} = 1.6, m_{11} = 0,$$
$$m_{12} = 0, m_{20} = -0.2, m_{21} = 0, m_{22} = 0.$$

  We assume a diffuse prior distribution for the variance-covariance matrix Σ_i. Under the assumption of noninformative priors, we set $A_0 = A_1 = \text{diag}(1,1), f_0 = f_1 = 2,$

where diag(a_i) is the $n \times n$ diagonal matrix with a_i elements. This assumption is repeated in the following analysis.

Figure 1 shows the contour plot of the joint distribution of the prior information of effectiveness and cost for each treatment, and the contour plot of the joint distribution of the prior incremental effectiveness and cost between treatments.

The posterior distribution was simulated using WinBUGS (Spiegelhalter, Thomas and Best, 2000b). A total of 100000 iterations were carried out (after a burn-in period of 50000 simulations). Convergence was evaluated for all parameters using several tests provided within the WinBUGS Convergence Diagnostics and Output Analysis software (CODA). The constant m_{00} is not required to ensure convergence.

Table 2 shows the posterior analysis for the independence case. The posterior incremental effectiveness is estimated at -0.02928 units with a standard deviation of 1.328. The incremental cost is estimated at 0.162 units.

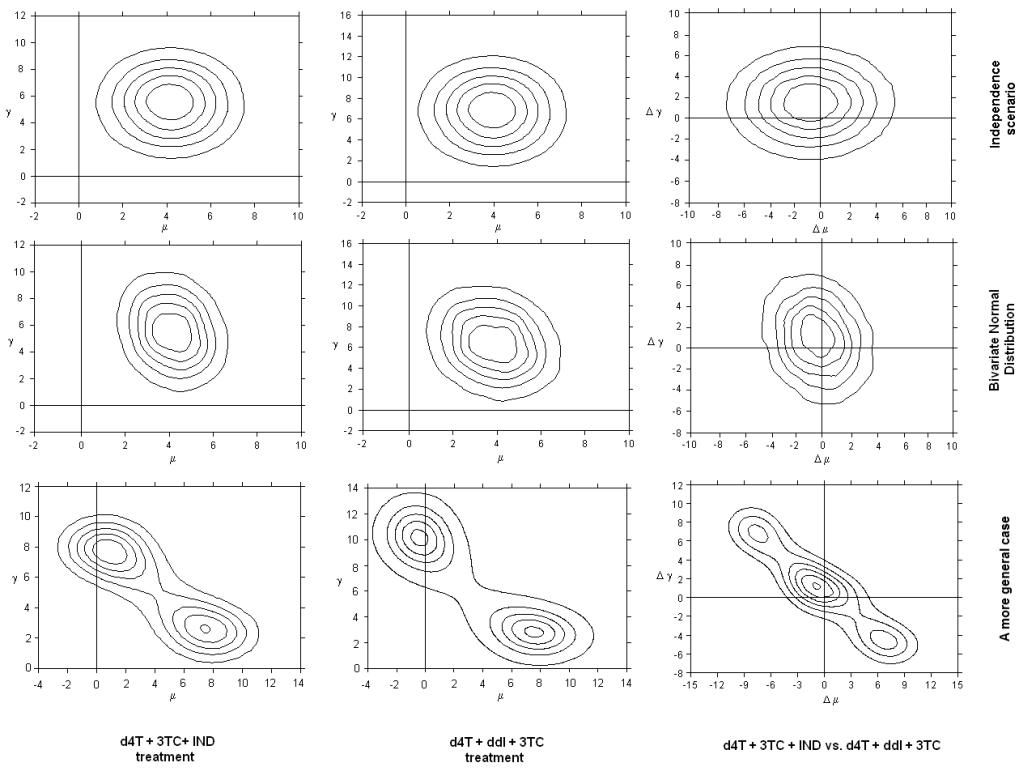

Figure 1: *Contour plots of the joint prior distribution of* μ *and* γ.

Table 2: *Posterior moments: mean and standard deviation.*

	Independence	Bivariate-Normal distribution	Bimodal case
μ_0	4.540 (0.7911)	4.45 (0.7862)	3.359 (0.777)
γ_0	7.137 (0.09542)	7.137 (0.0952)	7.128 (0.09661)
μ_1	4.507 (1.069)	4.422 (1.06)	2.152 (0.8683)
γ_1	7.302 (0.1784)	7.301 (0.1784)	7.293 (0.1824)
$\Delta\mu$	-0.02956 (1.328)	-0.02397 (1.318)	-1.207 (1.165)
$\Delta\gamma$	0.1628 (0.2028)	0.1613 (0.2023)	0.1643 (0.2063)

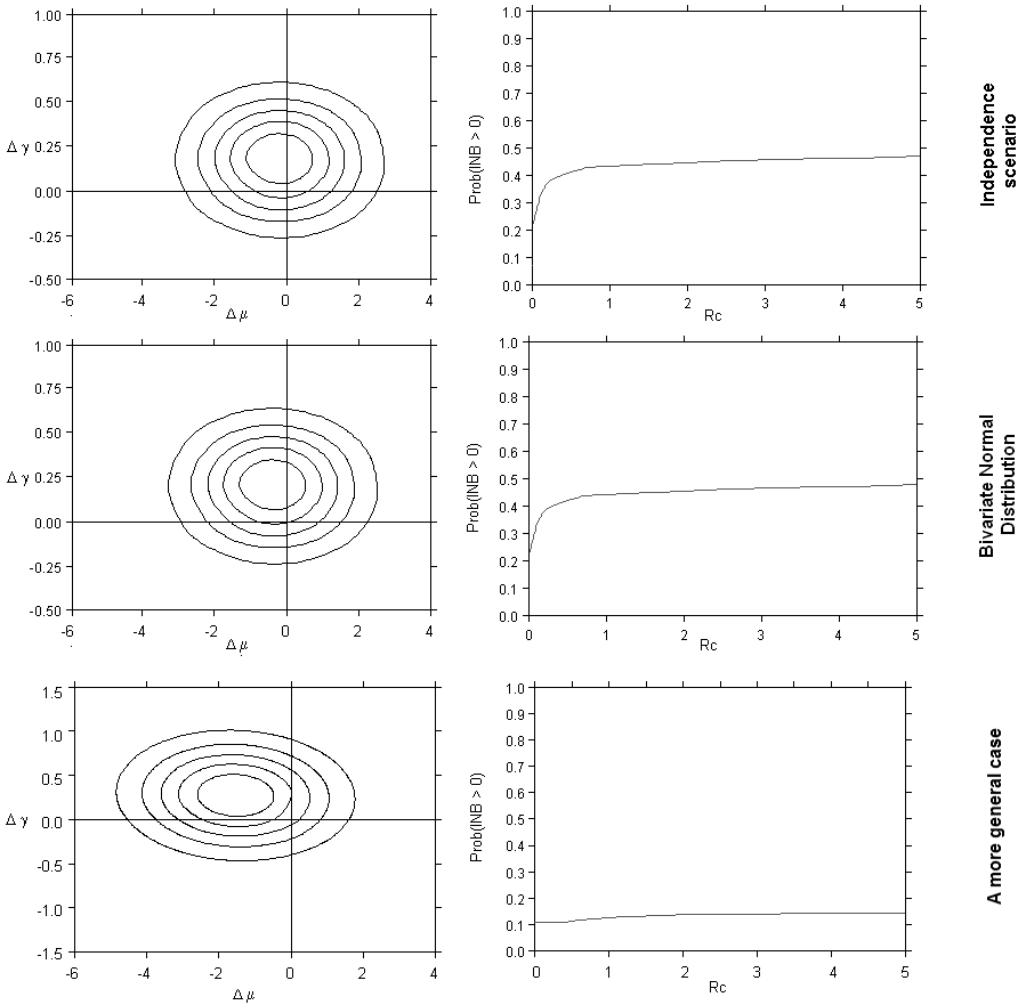

Figure 2: *Contour plots of the joint posterior distribution of μ and γ and the cost-effectiveness acceptability curves.*

Figure 2 shows the cost-effectiveness plane and the cost-effectiveness acceptability curve. It is apparent that the treatment (d4T + ddl + IND) will never be preferable to the treatment (d4T + 3TC + IND), as the probability of a positive INB is always below 50%.

- Bivariate normal distribution:

  The previous analysis was repeated under the assumption of a correlation between cost and effectiveness for each treatment. We asked the experts to assess the correlation directly by specifying a value between -1 and 1. Although many researchers have suggested that the direct assessment of moments is a poor method of quantifying opinion, Clemen, Fischer and Winkler (2000) found that this method performed best for eliciting a correlation. The experts considered a prior correlation of $\rho = -0.2$. By incorporating this information into the prior information described in the previous subsection, we calculated the following prior parameters:

$$m_{01} = 1.6146, m_{02} = -0.1302, m_{10} = 2.4306, m_{11} = -0.0694,$$
$$m_{12} = 0, m_{20} = -0.2315, m_{21} = 0, m_{22} = 0,$$

  for the (d4T + 3TC + IND) treatment, and

$$m_{01} = 1.2108, m_{02} = -0.0833, m_{10} = 1.9829, m_{11} = -0.0527,$$
$$m_{12} = 0, m_{20} = -0.2083, m_{21} = 0, m_{22} = 0,$$

  for the (d4T + ddl + IND) treatment.

  Figure 1 includes the joint distribution of the prior information on the effectiveness and cost for each treatment, and the joint distribution of the prior incremental effectiveness and cost between treatments. It also shows the negative correlation between effectiveness and cost. Figure 2 shows the cost-effectiveness plane and the cost-effectiveness acceptability curve. The results are similar to those reached in the independence case. The treatment (d4T + 3TC + IND) is always preferred for any willingness to pay.

- A more general case:

  This example shows a prior bimodal density for the effectiveness and cost of each treatment. The experts agreed that effectiveness and cost depend on the presence of complications during the treatment, mainly due to the existence of concomitant illnesses. In this paper, the dependence between effectiveness and cost is specified by a conditional probability among the elicitation variables of interest. We asked the experts for the conditional median and first and third quartiles to elicit the mean and the variance of the normal distributions.

If there were no complications during the study, the mean effectiveness of the (d4T + 3TC + IND) treatment would be close to 8 units, and the mean cost would be about 2000 euros. The mean effectiveness decreases to 1, and the mean cost increases to 8000 euros with the presence of complications.

The presence of complications has more costly consequences for the (d4T + ddl + 3TC) treatment. The mean cost increases to 10000 euros and the mean effectiveness is reduced to 0. Under favourable conditions, the mean effectiveness is about 8 units and the mean cost is about 3000 euros.

This prior information was elicited through a conditionally specified prior distribution, compiling information about the conditional moments. Descriptions of the conditions were given to the experts in written form. An example of a verbal statement of a conditional event (Garthwaite and Al-Awadhi, 2001) is

- Suppose that a large number of patients are examined and their average cost is 2000 euros. What is your median estimate of their effectiveness?
- Consider the situation in which we know that your median value is true. In the light of this, assess your quartiles for effectiveness.

Table 3 shows the conditional moments employed in the elicitation process.

Table 3: *Prior conditional moments.*

Moment	Condition	(d4T+3TC+IND)	(d4T+ddl+IND)
$E(\mu\|\gamma)$	$\gamma = 2$	8	8
	$\gamma = 5$	4	3.5
	$\gamma = 8$	2	0
$Var(\mu\|\gamma)$	$\gamma = 2$	2.5	2.5
	$\gamma = 5$	2.25	2
	$\gamma = 8$	1.75	1.5
$E(\gamma\|\mu)$	$\mu = 0$	8	10
	$\mu = 4$	5	6
	$\mu = 8$	2	3
$Var(\gamma\|\mu)$	$\mu = 0$	6	4.5
	$\mu = 4$	4.5	7.5
	$\mu = 8$	2.5	7

By using this prior information we can calculate the values of the hyperparameters, applying them to the optimization problem explained in the previous section:

$$m_{01} = 9.3931, m_{02} = -0.6198, m_{10} = 8.1442, m_{11} = -1.8114,$$
$$m_{12} = 0.1012, m_{20} = -0.5241, m_{21} = 0.1412, m_{22} = -0.0147$$

for the (d4T + 3TC + IND) treatment, and

$$m_{01} = 7.5074, m_{02} = -0.4014, m_{10} = 67.0746, m_{11} = -1.0390,$$
$$m_{12} = 0.0231, m_{20} = -0.4792, m_{21} = 0.1209, m_{22} = -0.0165$$

for the (d4T + ddl + IND) treatment.

Figure 1 shows the joint distribution of the prior information on the effectiveness and cost of each treatment, together with the joint distribution of the prior incremental effectiveness and cost between treatments. There was found to be a bimodal joint distribution for cost and effectiveness. This joint distribution, and the marginal distributions of effectiveness and cost were shown to the experts to assess the adequacy of the elicitation.

It is important to note that the mean of the marginal distributions of effectiveness and cost for both treatments coincides with the prior mean elicited in the "independent" section. However, this more general model opens up a wide range of possibilities for incorporating different prior beliefs far removed from those of the conventional bivariate normal distribution.

Figure 2 shows the measures used to take decisions. The analysis shows a preference for the treatment (d4T + 3TC + IND) for all the scenarios. In fact, the CEAC is always lower than the critical value 50%. It is important to point out that, although we have considered similar prior means of effectiveness and costs, the uncertainty about the right decision is different for the independence scenario, the bivariate normal distribution and the bimodal prior distribution. If we considered a willingness to pay of 5 euros, the probability of preferring the (d4T + 3TC + IND) for the first two scenarios is only 52%. This probability increases to 85% for the more general case. This is due to the fact that the latter model includes in the analysis the prior information that any complication arising during the treatment would have more important consequences with the (d4T + ddl + 3TC) treatment than with the control treatment.

4 Conclusions and discussion

The Bayesian approach allows the incorporation of prior information. In a fully Bayesian analysis, the procedures used to elicit expert opinion are an active research issue. This paper studies the use of a general family of prior distributions for the mean of the effectiveness and cost. In particular, we assume that the conditional density of the mean effectiveness for a given mean cost and the conditional density of the mean cost for a given mean effectiveness are both normal.

The improvement gained over the use of conditionally specified priors is the wide range of prior information that may be elicited. Prior information from more than one source, or different structures of effectiveness and costs depending on whether complications occur, are some cases whereby a conventional bivariate prior distribution is not enough to specify the prior information. Conventional cases, such as the independence case and bivariate prior information, are included as particular cases of this more general analysis. The posterior distribution is easily simulated using MCMC techniques (Gelman, Carlin, Stern and Rubin, 1995; Gilks, Richardson and Spiegelhalter, 1996; Gamerman and Lopes, 2006).

The practical application shows the sensitivity of the results to the prior distribution. The more general case, in which experts provide a bimodal prior distribution for mean effectiveness and mean cost, the probability of preference for the conventional treatment (d4T + 3TC + IND) varies from 85% to 89% in a willingness to pay range of $R_c \in (0,5)$. For the more conventional prior structure, bivariate or independent normal distributions, this probability varies from 65% to 52% for the same range of R_c.

However, this methodology also present some disadvantages. Psychological research has shown that conditional assessments can be affected by biases such as conservatism (Edwards and Phillips, 1964) and, intuitively, making assessments conditionally on hypothetical data is a more difficult task than making unconditional assessments. Besides, the large number of parameters present in high-dimensional conditionally specified priors is the source of their flexibility but, in practice, poses elicitation problems. In this context, the sensitivity analysis may play a crucial role (Stevens and O'Hagan, 2002). In our opinion, conditionally specified priors are not a panacea but certainly, for many classical data analysis situations, they offer a manageable and more flexible alternative to the usual, rather restrictive, priors.

Acknowledgments

We wish to thank the referees for their comments. Research partially funded by SEJ2006-12685 (MEC, Spain) and ECO2009-14152 (Ministerio de Ciencia e Innovación, Spain). We thank Xavier Badía for allowing us to use his data archive. All responsibility for the further analysis of the data, of course, is ours alone.

References

Arnold, B., Castillo, E. and Sarabia, J. (1998). Bayesian analysis for classical distributions using conditionally specified priors. *Sankhya-The Indian Journal of Statistics*, 60, 228-245.

Arnold, B., Castillo, E. and Sarabia, J. M. (1999). *C*onditional Specification of Statistical Models. Springer-Verlag, New York.

Arnold, B., Castillo, E. and Sarabia, J. (2001a). Bayesian inference using conditionally specified priors. *Handbook of Applied Econometrics and Statistical Inference*. Marcel Dekker, New York.

Arnold, B., Castillo, E. and Sarabia, J. (2001b). Conditionally specified distributions: an introduction. *Statistical Science*, 16, 249-274.

Arnold, B., Castillo, E., Sarabia, J. and González-Vega, L. (2000). Multiple modes in densitites with normal conditionals. *Statistics and Probability Letters*, 49, 355-363.

Baranski, J. V. and Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception and Psychophysics*, 55, 412-428.

Berger, J. O. (1994). An overview of robust Bayesian analysis (with discussion). *Test*, 3, 5-124.

Briggs, A. H. (1999). A Bayesian approach to stochastic cost-effectiveness analysis. *Health Economics*, 8, 257-261.

Brooks, R. and EuroQol group (1996). EuroQol: The current state of play. *Health Policy*, 37, 53-72.

Castillo, E. and Galambos, J. (1989). Conditional distributions and the bivariate normal distribution. *Metrika*, 36, 209-214.

Chesley, G. R. (1978). Subjective probability elicitation techniques: A performance comparison. *Journal of Accounting Research*, 16, 225-241.

Clemen, R. T., Fischer, G. W. and Winkler, L. R. (2000). Assessing dependence: Some experimental results. *Management Science*, 46, 1100-1115.

Cooke, R. M. and Slijkhuis, K. A. (2003). *Expert judgement in the uncertainty analysis of dike ring failure frequency.* In W. R. Blischke and D. N. Prabhakar Murthy (Eds.). *Case Studies in Reliability and Maintenance* (331-352). Wiley, Chichester.

Denham, R. and Mengersen, K. (2007). Geographically assisted elicitation of expert opinion for regression models. Bayesian Analysis. *Journal of the Royal Statistical Society, B*, 2, 99-136.

Diebolt, J. and Robert, C. P. (1994). Estimation of Finite Mixture Distributions through Bayesian Sampling. *Journal of the Royal Statistical Society, B*, 56, 363-375.

Edwards, W. and Phillips, L. D. (1964). *Man as transducer for probabilities in Bayesian command and control systems.* In G. L. Bryan and M. W. Shelley (Eds.). *Human Judgements and Optimality*. Wiley, New York.

Fenwick, E., Palmer, S., Claxton, K., Sculpher, M., Abrams, K. and Sutton, A. (2006). An iterative Bayesian approach to health technology assessment: application to a policy of preoperative optimization for patients undergoing major elective surgery. *Medical Decision Making*, 26, 480-496.

Fryback, D., Chinnis, J. and Ulvila, J. (2001). Bayesian cost-effectiveness analysis. An example using the GUSTO trial. *International Journal of Technology Assessment in Health Care*, 17, 83-97.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall, London.

Garthwaite, P. H. and Al-Awadhi, S. A. (2001). Non-conjugate prior distribution assessment for multivariate normal sampling. *Journal of the Royal Statistical Society, Series B*, 63, 95-110.

Garthwaite, P. H., Kadane, J. B. and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100, 680-701.

Gelman, A. and Meng, X. (1991). A note on bivariate distributions that are conditionally normal. *American Statistics*, 45, 125-126.

Gelman, R., Carlin, J., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Heitjan, D. (1997). Bayesian interim analysis of phase II cancer trials. *Statistics in Medicine*, 16, 1791-1802.

Heitjan, D., Moskowitz, A. and William, W. (1999). Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Economics*, 8, 191-201.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, 70, 271-294.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S. and Peters, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, 75, 845-854.

Kadane, J. B. and Wolfson, L. J. (1998). Experiences in elicitation. *The Statistician*, 47, 3-19.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.

Kennedy, M., Anderson, C., O'Hagan, A., Lomas, M., Woodward, I., Gosling, J. P. and Heinemeyer, A. (2008). Quantifying uncertainty in the biospheric carbon flux for England and Wales. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 109-135.

Laska, E., Meisner, M. and Siegel, C. (1997). Statistical inference for cost-effectiveness ratios. *Health Economics*, 6, 229-242.

Leal, J., Wordsworth, S., Legood, R. and Blair, E. (2007). Eliciting expert opinion for economic models: An applied example. *Value in Health*, 10, 195-203.

Löthgren, M. and Zethraeus, N. (2000). Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Economics*, 9, 623-630.

Malakoff, D. (1999). Bayes offers a new way to make sense of numbers. *Science*, 286, 1460-1464.

McIntosh, M. W., Ramsey, S. D., Berry, K. and Urban, N. (2001). Parameter solicitation for planning cost effectiveness studies with dichotomous outcomes. *Health Economics*, 10, 53-66.

Negrín, M. and Vázquez-Polo, F. J. (2006). Bayesian cost-effectiveness analysis with two measures of effectiveness: the cost-effectiveness acceptability plane. *Health Economics*, 15, 363-372.

O'Hagan, A. and Stevens, J. (2001). A framework for cost-effectiveness analysis from clinical trial data. *Health Economics*, 10, 303-315.

O'Hagan, A., Stevens, J. and Montmartin, J. (2001). Bayesian cost-effectiveness analysis from clinical trial data. *Statistics in Medicine*, 20, 733-753.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, J., Oakley, J. and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley and Sons, Chicester.

Pinto, J., López, C., Badía, X., Corna, A. and Benavides, A. (2000). Análisis coste-efectividad del tratamiento antirretroviral de gran actividad en pacientes infectados por el VIH asintomáticos. *Medicina Clínica*, 114, 62-67.

Saltelli, A., Chan, K. and Scott, E. M. (2000). *Sensitivity Analysis*. Wiley, Chichester.

Sarabia, J. M. and Gómez-Déniz, E. (2008). Construction of multivariate distributions: a review of some recent results. *SORT*, 32, 3-36.

Sarabia, J. M., Castillo, E., Gómez-Déniz, E. and Vázquez-Polo, F. J. (2005). A class of conjugate priors for log-normal claims based on conditional specification. *Journal of Risk and Insurance*, 72, 479-495.

Sarmanov, O. V. (1966). Generalized normal correlation and two-dimensional Frechet classes. *Doklady (Soviet Mathemathics)*, 168, 596-599.

Smith, M. K. and Marshall, S. (2006) A Bayesian design and analysis for dose-response using informative prior information. *Journal of Biopharmaceutical Statistics*, 16, 695-709.

Spiegelhalter, D., Feedman, L. and Parmar, M. (1994). Bayesian approaches to randomized trials (with discussion). *Journal of the Royal Statistical Society, Series A*, 157, 357-416.

Spiegelhalter, D. J., Myles, J. P., Jones, D. R. and Abrams, K. R. (2000a). Bayesian methods in health technology assessment: a review. *Health Technol Assess*, 4.

Spiegelhalter, D., Thomas, A. and Best, N. (2000b). WinBUGS version 1.3. User manual, Biostatistics Unit, Cambridge Medical Research Council.

Staël von Holstein, C. A. S. (1970). Measurement of subjective probability. *Acta Psychologica*, 34, 146-159.

Stevens, J. and O'Hagan, A. (2002). Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data. *International Journal of Technology Assessment in Health Care*, 18, 782-790.

Stinnett, A. and Mullahy, J. (1998). Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making*, 18, S65-S80.

Tambour, M., Zethraeus, N. and Johannesson, M. (1998). A note on confidence intervals in cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care*, 14, 467-471.

Ting-Lee, M. L. (1996). Properties and applications of the Sarmanov family of bivariate distributions. *Communications Statistics: Theory and Methods*, 25, 1207-1222.

Vázquez-Polo, F. J. and Negrín, M. (2004). Incorporating patients' characteristics in cost-effectiveness studies with clinical trial data: a flexible Bayesian approach. *SORT*, 28, 87-108.

Viviane, A. and Barkun, N. A. (2008). Estimates of costs of hospital stay for variceal and nonvariceal upper gastrointestinal bleeding in the United States. *Value in Health*, 11, 1-3.

Willan, A. and O'Brien, B. (1996). Confidence intervals for cost-effectiveness ratios: an application of Fielle's theorem. *Health Economics*, 5, 297-305.

Winkler, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, 62, 776-800.

Winkler, R. L. (2001). Why Bayesian analysis hasn't caught on in healthcare decision making. *International Journal of Technology Assessment in Health Care*, 17, 56-66.

Yaniv, I., Yates, J. F. and Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, 110, 611-617.

Yates, J. F. (1990). *Judgment and decision making*. Prentice-Hall, Englewood Cliffs.