COLLANA DEL
DIPARTIMENTO DI ECONOMIA

A BAYESIAN MODEL AVERAGING APPROACH WITH NON-INFORMATIVE PRIORS FOR COST-EFFECTIVENESS ANALYSES IN HEALTH ECONOMICS

Caterina Conigliani

Working Paper n° 94, 2008

- I “Working Papers” del Dipartimento di Economia svolgono la funzione di divulgare tempestivamente, in forma definitiva o provvisoria, i risultati di ricerche scientifiche originali. La loro pubblicazione è soggetta all’approvazione del Comitato Scientifico.

- Per ciascuna pubblicazione vengono soddisfatti gli obblighi previsti dall’art. 1 del D.L.L. 31.8.1945, n. 660 e successive modifiche.

- Copie della presente pubblicazione possono essere richieste alla Redazione.

REDAZIONE:
Dipartimento di Economia
Università degli Studi Roma Tre
Via Silvio D'Amico, 77 - 00145 Roma
Tel. 0039-06-57335655 fax 0039-06-57335771
E-mail: dip_eco@uniroma3.it

ROMA TRE
UNIVERSITÀ DEGLI STUDI

DIPARTIMENTO DI ECONOMIA

A BAYESIAN MODEL AVERAGING APPROACH WITH NON-INFORMATIVE PRIORS FOR COST-EFFECTIVENESS ANALYSES IN HEALTH ECONOMICS

Caterina Conigliani

** Dipartimento di Economia, Università degli Studi "Roma Tre"*

This work is dedicated
to Dave Laws
a lovely statistician
a convinced Bayesian
and a great friend

A Bayesian model averaging approach with non-informative priors for cost-effectiveness analyses in health economics

Caterina Conigliani

Keywords: Bayesian model averaging, Cost data, Health economics, MCMC, Non-informative priors, Path Sampling, Sensitivity analysis

∂ EL classification: C11; C15.

Abstract

We consider the problem of assessing new and existing technologies for their cost-effectiveness in the case where data on both costs and effects are available from a clinical trial, and we address it by means of the cost-effectiveness acceptability curve. The main difficulty in these analyses is that cost data usually exhibit highly skew and heavy-tailed distributions, so that it can be extremely difficult to produce realistic probabilistic models for the underlying population distribution, and in particular to model accurately the tail of the distribution, which is highly influential in estimating the population mean. Here, in order to integrate the uncertainty about the model into the analysis of cost data and into cost-effectiveness analyses, we consider an approach based on Bayesian model averaging in the particular case of weak prior informations about the unknown parameters of the different models involved in the procedure. The main consequence of this assumption is that the marginal densities required by Bayesian model averaging are undetermined. However in accordance with the theory of partial Bayes factors and in particular of fractional Bayes factors, we suggest replacing each marginal density with a ratio of integrals, that can be efficiently computed via Path Sampling. The results in terms of cost-effectiveness are compared with those obtained with a semi-parametric approach that does not require any assumption about the distribution of costs.

1 Introduction

The increasing burden on the budgets of health care providers has resulted in considerable interest in assessing new and existing technologies for their clinical effectiveness and cost-effectiveness.

Suppose that we intend to compare two health care technologies T_1 and T_2 in a randomised controlled trial, where data consist of the effect e_{ij} and the cost c_{ij} of treatment i on patient j ($i = 1, 2$; $j = 1, 2, ..., n_i$).

In order to assess if T_2 is more cost-effective than T_1, we need to compare the expected effects μ_1 and μ_2 as well as the expected costs γ_1 and γ_2. In particular, let $\Delta_e = \mu_2 - \mu_1$ and $\Delta_c = \gamma_2 - \gamma_1$ be the effect and cost differentials. Moreover, let K be a decision-maker's *willingness to pay* coefficient, that is the units of money a decision maker is prepared to pay to obtain one unit of effectiveness.

The primary measure of cost-effectiveness of T_2 relative to T_1 is usually considered to be the *net monetary benefit* $K\Delta_e - \Delta_c$ (O'Hagan *et al.*, 2000): T_2 is cost-effective relative to T_1 if $K\Delta_e - \Delta_c > 0$, *i.e.* if in the plane of possible pairs of values of the population mean increments of effect and cost, (Δ_e, Δ_c) is below a sloping line of gradient K. This is usually referred as the *Net Benefit approach* (Stinnett and Mullahy, 1998), and inference about the net monetary benefit is generally presented by means of a Cost-Effectiveness Acceptability Curve (CEAC), that plots the probability $Q(K) = P(K\Delta_e - \Delta_c > 0)$ that

the net benefit is positive against the coefficient K (van Hout *et al.*, 1994), which is rarely unambiguously determined in practice. In this sense, a Bayesian approach is particularly natural, since no such probability exists or has any meaning in frequentist statistics (O'Hagan *et al.*, 2000).

Clearly these cost-effectiveness analyses of clinical trial data rely on statistical models which describe the distribution of costs and effects and their interrelation across individuals in the trial, which are rather difficult to determine, mainly because cost data obtained for individual patients in health economic studies typically exhibit highly skew and heavy tailed distributions. In fact, as discussed in O'Hagan and Stevens (2002, 2003), non-parametric methods, such as those based on the asymptotic normality of the sample mean or nonparametric bootstrapping, may be inefficient for analising such data and their justification breaks down in small samples. See Dinh and Zhou (2006) for some recent developments on such methods. On the other hand, parametric modelling may lead to more efficient inference (see, among others, Al and van Hout, 2000, O'Hagan and Stevens, 2001, 2002, Fryback *et al.*, 2001) but is dependent on the population distribution matching the model adequately. The main difficulty in this sense, as pointed out for instance in Nixon and Thompson (2004) and Thompson and Nixon (2005), is that the high skewness and kurtosis usually found in cost data imply that the population mean can be very sensitive to the tail of the distribution, that might be difficult to model accurately. One consequence of this is that parametric models that fit the data equally well can produce very different answers; conversely, in some cases models that fit badly can give similar inferences to those that fit well. For these reasons, Thompson and Nixon (2005) recommend that the sensitivity of conclusions to the choice of the model is always investigated, so that model uncertainty becomes a crucial aspect of analysing cost data. A different proposal for overcoming these difficulties can be found in Conigliani and Tancredi (2005a), that suggested to model the bulk of the data and the tails separately, with a distribution composed of a piecewise constant density up to an unknown endpoint and a generalised Pareto distribution (GPD) for the remaining tail data; this semi-parametric model, that is extremely flexible and able to fit data set with very different shapes, has been applied to cost-effectiveness analyses in the simple case where effects are measured as binary outcomes in Conigliani and Tancredi (2005b), and to more general settings in Conigliani and Tancredi (2008), where the results of the semi-parametric model are compared with those obtained with Bayesian model averaging (BMA). Note that an approach based on BMA in this setting is somehow in the spirit of the sensitivity analyses advocated by Thompson and Nixon (2005). In fact, it requires the specification of a set of plausible models for costs, but instead of studying how the conclusions change with the different models, it takes into account the inferences obtained with all the models that have a non-zero posterior probability. Obviously the main difficulty of this approach is the specification of the set of plausible models, in the sense that it should include distributions with a

wide range of shapes both for the bulk of data and for the tail. But another difficulty is represented by the fact that Bayesian model averaging requires proper prior distributions for the unknown parameters of the various models, even when there is not enough prior knowledge to elicit them.

Here, in order to focus the attention on the distribution of costs, we find convenient to write the distribution for a single observation $x_{ij} = (e_{ij}, c_{ij})'$ under treatment T_i as

$$f\left(x_{ij}\,|\theta_i, \phi_i\right) = f\left(c_{ij}\,|\theta_i\right) f\left(e_{ij}\,|c_{ij}, \phi_i\right)$$

where $f\left(c_{ij}\,|\theta_i\right)$ is the unconditional distribution for the cost of patient j under T_i and $f\left(e_{ij}\,|c_{ij}, \phi_i\right)$ is the conditional distribution for the effect on patient j under T_i given the cost c_{ij}, that we assume independent on the parameter θ_i of the distribution of costs. And in order to integrate the uncertainty about the model for costs into cost-effectiveness analyses, we consider an approach based on Bayesian model averaging in the particular case of weak prior informations for the unknown parameters of the different models. This is presented in details in Section 2, together with the problems caused by the assumption of non-informative priors, namely the fact that the marginal densities required by BMA are undetermined. In the same section we recall the theory of partial Bayes factors and in particular of fractional Bayes factors, and we suggest computing the required posterior model probabilities by replacing each (undetermined) marginal density with a ratio of integrals. In Section 3 we revise some of the numerical methods usually applied in Bayesian statistics to obtain inferences based on integrals, and focus our attention on Path sampling, a particularly flexible and efficient simulation method introduced by Gelman and Meng (1994, 1998) for the direct computation of ratios of marginal densities. In Section 4 the proposed methodology is compared with the semi-parametric approach of Conigliani and Tancredi (2005a) in an empirical context. A few concluding remarks are presented in the final section.

2 Analysing cost data with Bayesian model averaging with weak prior informations

Suppose that under each treatment group, instead of choosing a single parametric model for cost data, we specify a set of plausible models $\mathcal{M} = \{M_1, M_2, ..., M_k\}$. In particular, we assume that $\mathcal{M}$ is made of the log-normal, gamma, Weibull, log-logistic, generalised Pareto and inverse Gaussian; the corresponding probability density function and main summaries are shown in Table 1. Note that these distributions are always positive and skewed to the right, and offer a range of different tail behaviours. Moreover, we assume that all six distributions have finite first and second moment (which requires a constraint for the shape parameter of the log-logistic and the generalised Pareto), so that they can be re-parametrized

Table 1: Single parametric models included in the BMA procedure

	Pdf	Mean	Variance
Log − normal	$p(x\mid\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{[\log(x)-\mu]^2}{2\sigma^2}}$	$e^{\mu+\frac{\sigma^2}{2}}$	$e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$
Gamma	$p(x\mid\nu,\lambda) = \frac{x^{\nu-1}\lambda^{\nu}}{\nu(\nu)} e^{-x\lambda}$	$\frac{\nu}{\lambda}$	$\frac{\nu}{\lambda^2}$
Weibull	$p(x\mid\beta,\delta) = \frac{\beta x^{\beta-1} e^{-(x/\delta)^{\beta}}}{\delta^{\beta}}$	$\frac{\delta}{\beta}\Gamma\left(\frac{1}{\beta}\right)$	$\delta^2\left[\Gamma\left(\frac{2}{\beta}+1\right) - \Gamma\left(\frac{1}{\beta}+1\right)^2\right]$
Log − logistic	$p(x\mid\beta,\rho) = \frac{\rho(x/\beta)^{\rho}}{x[1+(x/\beta)^{\rho})]^2}$	$\frac{\pi\beta}{\rho}\csc\left(\frac{\pi}{\rho}\right), \quad \rho > 1$	$\frac{2\pi\beta^2}{\rho}\csc\left(\frac{2\pi}{\rho}\right) - \frac{\pi^2\beta^2}{\rho^2}\left[\csc\left(\frac{\pi}{\rho}\right)\right]^2, \quad \rho > 2$
GPD	$p(x\mid\sigma,\xi) = \frac{1}{\sigma}\left(1+\frac{x\xi}{\sigma}\right)^{-1/\xi-1}$	$\frac{\sigma}{(1-\xi)}, \quad \xi < 1$	$\frac{\sigma^2}{(1-\xi)^2(1-2\xi)}, \quad \xi < \frac{1}{2}$
InverseGaussian	$p(x\mid\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} exp\left[-\frac{\lambda(x-\mu)^2}{2x\mu^2}\right]$	μ	$\frac{\mu^3}{\lambda}$

in terms of the mean cost γ and the variance σ^2. Then the posterior marginal distribution of γ can be obtained by Bayesian model averaging (Hoeting *et al.*, 1999) as a mixture of its posterior marginal distributions under each of the models in $\mathcal{M}$, with weights given by the corresponding posterior model probabilities.

Formally, let $f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)$ and $\pi_l(\gamma,\sigma^2)$ be respectively the distribution of the cost data and the prior distribution of the parameters under model M_l in $\mathcal{M}$. Moreover, let $\pi(M_l)$ be the prior model probability of M_l, such that $\sum_{l=1}^{k}\pi(M_l)=1$. Then according to BMA the posterior marginal distribution of γ can be written as

$$\pi(\gamma|c_1,...,c_n) \quad = \quad \sum_{l=1}^{k}\left[\int \pi_l(\gamma,\sigma^2|c_1,...,c_n)d\sigma^2\right]\pi(M_l|c_1,...,c_n) \tag{1}$$

where the posterior distribution of the parameters under M_l and the posterior model probability of M_l, that need to be substituted in (1), can be obtained by means of Bayes' theorem as

$$\pi_l(\gamma,\sigma^2|c_1,...,c_n) = \frac{f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)}{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2} \tag{2}$$

and

$$\pi(M_l|c_1,...,c_n) = \frac{\pi(M_l)\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}{\sum_{l=1}^{k}\pi(M_l)\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2} \tag{3}$$

respectively.

Note that this is more or less the setting of Conigliani and Tancredi (2008); in particular, they suggest to re-parametrize all models in $\mathcal{M}$ in terms of the mean cost and the coefficient of variation (*i.e.* the ratio of the standard deviation to the mean), so that the unknown parameters have a clear meaning and the same prior distribution can be introduced under the various models in $\mathcal{M}$.

However Bayesian model averaging requires proper prior distributions for the unknown parameters under the various models in $\mathcal{M}$, even when there is not enough prior knowledge to elicit them. In fact,

if the prior for γ and σ^2 under model M_l is improper, *i.e.* defined only up to an arbitrary constant, then the marginal density $\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2$ is itself a multiple of such constant ($l=1,...,k$). Clearly this does not represent a problem for the computation of the posterior distribution of the parameters under model M_l, given by (2), since the same prior distribution, and therefore the same unknown constant, appears in the numerator and in the denominator of (2) and cancels out. Instead, the posterior model probability for M_l, given by (3), depends not only on $\pi_l(\gamma,\sigma^2)$, but also on the prior distribution for γ and σ^2 under the other models in $\mathcal{M}$. And, as pointed out for instance in Berger and Pericchi (1996), what appears to be the same parameter can have very different interpretations, and therefore different prior distributions, under different models. It follows that in (3) the different unknown constants do not cancel, and (3) cannot be computed.

Now, the fact that improper prior distributions result in undetermined marginal densities and represent a problem for computing posterior model probabilities, *i.e.* for Bayesian model selection, is a well known fact and several suggestions can be found in the literature to overcome this difficulty, most of which are based on the idea of training samples.

Formally, for each model M_l in $\mathcal{M}$, divide the cost data $c_1,...,c_n$ into two parts, $c_{(m)}$ and $c_{(n-m)}$, of size m and $n-m$ respectively, with $0<m<n$. First, subsample $c_{(m)}$ is used to update the prior $\pi_l\left(\gamma,\sigma^2\right)$ and obtain the posterior distribution $\pi_l\left(\gamma,\sigma^2\left|\,c_{(m)}\right.\right)$; in the second step, taking this as a prior distribution, the remaining data $c_{(n-m)}$ are used to compute the marginal density:

$$\int f_l\left(c_{(n-m)}\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2\left|c_{(m)}\right.)d\gamma d\sigma^2=\frac{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}{\int f_l\left(c_{(m)}\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2} \tag{4}$$

that clearly does not depend on an arbitrary constant if the prior distribution $\pi_l(\gamma,\sigma^2)$ is improper. Note that this is the idea that lead to the definition of partial Bayes factors. There is, however, a difficulty with the use of (4), namely the selection of the training sample $c_{(m)}$ from the data. To avoid the arbitrariness of choosing a particular training sample, O'Hagan (1995) suggested instead the use of a proportion $b=m/n$ of the data for training: if both m and n are large, the likelihood $f_l\left(c_{(m)}\left|\gamma,\sigma^2\right.\right)$ based only on the training sample $c_{(m)}$ will approximate to the full likelihood $f_i\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)$ raised to the power of b. By analogy with (4) is then:

$$\int f_l\left(c_{(n-m)}\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2\left|c_{(m)}\right.)d\gamma d\sigma^2\approx\frac{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)^b\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2} \tag{5}$$

which motivated the definition of the fractional Bayes factor (O'Hagan, 1995).

Now going back to the problem of deriving the posterior model probabilities (3) required by (1) when there is not enough prior knowledge to elicit proper prior distributions for γ and σ^2, the above

approximation could be used to replace each marginal density in (3), leading to

$$\pi(M_l|c_1,...,c_n) \approx \frac{\pi(M_l)\dfrac{\int f_l\left(c_i,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)^b\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}}{\sum_{l=1}^{k}\pi(M_l)\dfrac{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)^b\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}} \tag{6}$$

One last issue is worth a few considerations. The choice of the size of the training sample has been widely discussed in the literature; see for instance O'Hagan (1995, 1997). One simple and obvious guidance, that has proved to be reliable in a range of problems involving improper priors, is to consider the minimal training sample, *i.e.* the smallest sample size needed to update an improper prior so as to obtain a proper prior distribution. Here, if we consider an improper prior for γ and σ^2, a minimal training sample is any subset made of two observations, so that $b = 2/n$.

3 Computing posterior model probabilities via path sampling

Computing marginal densities, or equivalently normalising constants of probability models, is a fundamental computational problem for many statistical and scientific studies; for a review of the methods more widely used in Bayesian statistics see, for instance, Smith *et al.* (1985), Smith (1991), Tanner (1993), O'Hagan and Forster (2004), and the references therein; for comparisons of these methods in Bayesian model selection see, for example, Rosenkrantz (1992), Kass and Raftery (1995), Raftery (1996).

Approximations of marginal densities for well behaved problems of modest dimensionality can be obtained by a number of different procedures; these include methods of analytic approximation, such as Laplace's method (De Bruijn, 1961; Tierney and Kadane, 1986), and numerical integration procedures, such as adaptive Gaussian quadrature (Genz and Kass, 1993) and Bayesian quadrature (O'Hagan, 1991). For complex models, however, the only methods available are those based on simulations.

Simulation-based procedures include *Monte Carlo integration* (Hammersley and Handscomb, 1979) and *importance sampling* (Geweke, 1989); the idea is to learn about a complex probability distribution by simulating a set of random numbers from it, or from an auxiliary distribution that approximates it. In many fields, however, the complex probability systems encountered make these methods often unusable, and more advanced and typically more efficient simulation procedures are in common use. In particular, since the advent of *Markov chain Monte Carlo* methods (Gelfand and Smith, 1990, Tanner and Wong, 1987), several procedures have been proposed to obtain estimates of marginal densities by sampling from the posterior distribution of the parameters; work along this line includes Newton and Raftery (1994), Gelfand and Dey (1994), Chib (1995), Chib and Jeliazkov (2001).

Notice that the majority of the procedures we have referred to only produce approximations of one integral at a time. In various applications however, as in the case of computing posterior model probabilities via (5), the real interest is often not the magnitude of the integral, but rather ratios, or equivalently differences of the logarithms. Moreover such methods can be particularly unstable when the integrand is diffuse, as it is typically the case with the denominator of (5) for small b. It follows that in this setting procedures for the direct approximation of ratios of integrals, such as the *Bridge Sampling* of Meng and Wong (1996) and the *Path Sampling* of Gelman and Meng (1994, 1998), may be more appropriate.

A different approach that leads to the direct approximation of ratios of marginal densities is related to the specialized MCMC algorithms developed to handle problems involving inference about curves, surfaces or images, where the dimension of the object of inference is not fixed. Work along this line includes the product-space approach of Carlin and Chib (1995) and the *reversible jump* MCMC method of Green (1995). However both methods require the specification of proper priors for the unknown parameters, so that here we focus our attention on Path Sampling.

Thus, suppose we intend to compute the ratio of integrals (5) in order to obtain the posterior model probabilities (6), and let $\lambda(M_l)$ be the log of (5), *i.e.*

$$\lambda(M_l) = \log\left(\frac{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}{\int f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)^b\pi_l(\gamma,\sigma^2)d\gamma d\sigma^2}\right) \tag{7}$$

The fundamental idea underlying path sampling is that it is always possible to construct a continuous path $q_l\left(\gamma,\sigma^2\left|\,\omega\right.\right)$, $\omega\in\Omega$, with normalising constant $z_l\left(\omega\right)=\int q_l\left(\gamma,\sigma^2\left|\,\omega\right.\right)d\gamma d\sigma^2$, connecting the two functions that need to be integrated in (7). For instance in this setting it is quite obvious to choose

$$q_l\left(\gamma,\sigma^2\left|\,\omega\right.\right)=f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)^{\omega}\pi_l(\gamma,\sigma^2)$$

with $\omega\in[b,1]$, so that (7) can be rewritten as

$$\lambda(M_l)=\log\left(\frac{\int q_l\left(\gamma,\sigma^2\left|\,1\right.\right)d\gamma d\sigma^2}{\int q_l\left(\gamma,\sigma^2\left|\,b\right.\right)d\gamma d\sigma^2}\right)=\log\left(\frac{z_l(1)}{z_l(b)}\right). \tag{8}$$

Following Gelman and Meng (1998), to derive the basic identity for path sampling, let $p_l\left(\gamma,\sigma^2\left|\,\omega\right.\right)=q_l\left(\gamma,\sigma^2\left|\,\omega\right.\right)/z_l\left(\omega\right)$. Taking logarithms of $z_l\left(\omega\right)$ and differentiating with respect to ω yields

$$\frac{d}{d\omega}\log\left[z_l\left(\omega\right)\right]=\int p_l\left(\gamma,\sigma^2\left|\,\omega\right.\right)\frac{d}{d\omega}\log q_l\left(\gamma,\sigma^2\left|\,\omega\right.\right)d\gamma d\sigma^2=E_\omega\left[\log f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\right]$$

where E_ω denotes the expectation with respect to the sampling distribution $p_l\left(\gamma,\sigma^2\left|\,\omega\right.\right)$. It follows that (8) can be rewritten as

$$\lambda(M_l)=\int_b^1 E_\omega\left[\log f_l\left(c_1,...,c_n\left|\gamma,\sigma^2\right.\right)\right]d\omega, \tag{9}$$

which represents the key formula for path sampling: $\lambda(M_l)$ can be expressed as the integral over Ω of an expected value with respect to the conditional distribution $p_l\left(\gamma, \sigma^2 \,|\, \omega\right)$, of a function of the path.

Note that in order to evaluate (9), one must be able to compute the expected value $E_\omega\left[\log f_l\left(c_1, ..., c_n \,|\gamma, \sigma^2\right)\right]$ for given ω, and to carry out the integration with respect to ω. In particular, the latter is usually obtained by standard numerical integration, that requires the integrand to be evaluated at a set of fixed points ω_r $(r = 1, ..., N)$. Instead, the expected value is usually obtained by simulation, *i.e.* by sampling values $(\gamma_{sr}, \sigma^2_{sr})$ $(s = 1, ..., M)$ directly and independently from $q\left(\gamma, \sigma^2 \,|\, \omega_r\right)$ for given ω_r or, when this is not practical, with an iterative Markov chain Monte Carlo sampler. The corresponding estimate then has the form

$$\widehat{\lambda}(M_l) = \sum_{r=1}^{N} a_r \left\{ \frac{1}{M} \sum_{s=1}^{M} \left[\log f_l\left(c_1, ..., c_n \,|\gamma_{sr}, \sigma^2_{sr}\right)\right] \right\} \tag{10}$$

where $a_1, ..., a_N$ is a set of weights.

Notice that an alternative to numerical integration is to evaluate the integral with respect to ω by Monte Carlo importance sampling, by drawing $\omega_1, \omega_2, ..., \omega_N$ randomly from an arbitrary density $p(\omega)$. However this approach will typically be less efficient and more computationally demanding than numerical integration, that usually involves studying the function we are trying to integrate, and choosing the points ω_r in some sensible way. Moreover, recall that in our approach inference on the mean cost γ is based on the posterior marginal distribution (1), that depends on the posterior distributions of γ and σ^2 under the different models in $\mathcal{M}$ (as well as on the posterior model probabilities). And posterior summaries of $\pi_l(\gamma, \sigma^2 | c_1, ..., c_n)$ will typically require simulating from it, *i.e.* from $q_l\left(\gamma, \sigma^2 \,|\, \omega = 1\right)$. It follows that we find working with a fixed grid (such that $\omega_N = 1$) preferable.

The main difficulty in this approach is that the evaluation of $E_\omega\left[\log f_l\left(c_1, ..., c_n \,|\gamma, \sigma^2\right)\right]$ requires sampling from $q_l\left(\gamma, \sigma^2 \,|\, \omega_r\right)$ for values of ω_r that explore the whole interval $[b, 1]$, and $q_l\left(\gamma, \sigma^2 \,|\, \omega_r\right)$ will typically be diffuse for ω_r close to b. It follows that (10) can be rather unstable for small values of b. For this reason, we suggest computing (10) with an adaptive procedure of some kind, and in particular with an adaptive Metropolis-Hastings algorithm.

Consider first the problem of sampling from $q_l\left(\gamma, \sigma^2 \,|\, \omega_N = 1\right)$, *i.e.* from the posterior distribution of γ and σ^2 under model M_l; this is a standard computational problem and can be usually solved by proposing from any sensible distribution. Then consider the problem of sampling from $q_l\left(\gamma, \sigma^2 \,|\, \omega_{N-1}\right)$; now the target density is more diffuse than $q_l\left(\gamma, \sigma^2 \,|\, \omega_N = 1\right)$, but only slightly, so that useful indications about the proposal distribution can be obtained by looking at the values $(\gamma_{1N}, \sigma^2_{1N}), ..., (\gamma_{sN}, \sigma^2_{sN}), ..., (\gamma_{MN}, \sigma^2_{MN})$ simulated from $q_l\left(\gamma, \sigma^2 \,|\, \omega_N = 1\right)$. And so on: in order to sample from $q_l\left(\gamma, \sigma^2 \,|\, \omega_r\right)$, that is slightly more diffuse than $q_l\left(\gamma, \sigma^2 \,|\, \omega_{r+1}\right)$, we can adapt the proposal distribution in the Metropolis algorithm by

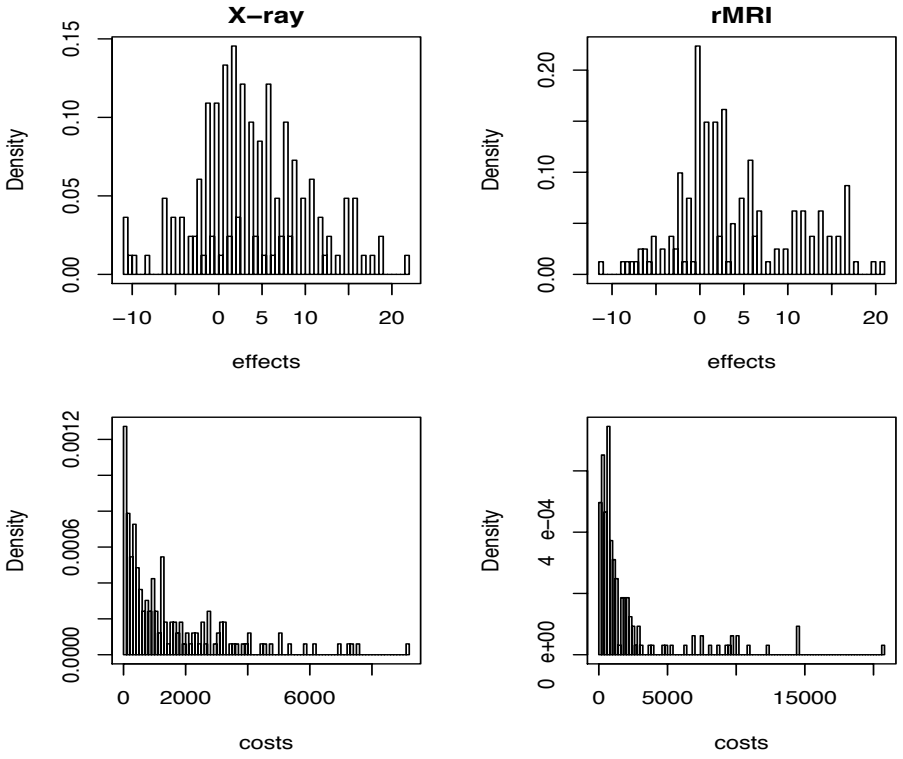

Figure 1: Data from the low back pain trial

looking at the values simulated from $q_l\left(\gamma, \sigma^2 \,|\, \omega_{r+1}\right)$ $(r = 1, ..., N-1)$. By doing so, as the target density becomes more diffuse, also the proposal distribution becomes more diffuse, and (10) does not suffer from stability problems for small values of b.

4 Analysis of the Low Back Pain Trial data

We present an example using a study on low back pain (Jarvik *et al.*, 2003). A total of 380 patients (out of which 328 were included in the health economic evaluation) were randomised in a 1:1 ratio to investigation by standard X-ray investigation and rapid magnetic resonance imaging (rMRI), and were followed for 12 months. Aim of the trial was to investigate whether rMRI would allow better diagnosis and treatment, or lead to unnecessary treatment without improvement in symptoms. The primary clinical endpoint was the change from baseline of the modified Roland back pain score (Patrick *et al.*, 1995), while the primary economic endpoint was the total health care cost (in US$).

The data are shown in Figure 1. Under both treatments the effects as well the marginal conditional

Table 2: Low back pain trial: sample descriptive statistics of X-ray costs and rMRI costs

	X ray	**rMRI**
sample size	166	162
mean	1515	2187
standard deviation	1747	3378
median	926	871
minimum	44	49
maximum	9111	20664
skewness $\left(\overline{\mu}_3/\sigma^3\right)$	1.8	2.7
kurtosis $\left(\overline{\mu}_4/\sigma^4\right)$	6.4	10.9

distributions of the effects are apparently well represented by a normal distribution, so that we assume

$$e_{ij}\,|c_{ij} \sim N\left(\beta_i + \delta_i c_{ij}, \sigma_i^2\right),$$

and the overall mean effect can be written as $\mu_i = \beta_i + \delta_i \gamma_i$. Instead, the distribution of costs is clearly highly skew and heavy-tailed, and this fact is confirmed also by the sample summaries shown in Table 2.

We now apply the approach of Section 2 and Section 3 to this data set, assuming for both treatment groups equal prior model probabilities for the six distributions in $\mathcal{M}$, and introducing the standard non-informative prior $\pi_l(\gamma, \sigma^2) \propto \dfrac{1}{\sigma^2}$ for the unknown parameters γ and σ^2 under each model in $\mathcal{M}$ $(l = 1, ..., k)$. In particular, we computed the ratios (5) required by the posterior model probabilities (6) via Path Sampling, applying the trapezoidal rule on a grid of N=10 points for integrating with respect to ω, and with the adaptive Metropolis-Hastings algorithm outlined in the previous section sampling $M = 20000$ values from each $q_l\left(\gamma, \sigma^2\,|\,\omega_r\right)$ $(r = 1, ..., N)$. As proposal distributions for γ and σ^2 we used two independent skew-t distributions, that turned out to be flexible and (when necessary) heavy tailed enough for this problem, allowing the adaptive procedure of Section 3 to work quite well as the target density $q_l\left(\gamma, \sigma^2\,|\,\omega_r\right)$ became more diffuse. The results are presented in Table 3 and Table 4 for the individual mean costs γ_1 and γ_2 respectively, and in Table 5 for the cost differential Δ_c and effect differential Δ_e.

Note that Table 3 and Table 4 also show the posterior summaries that we obtain if, instead of applying Bayesian model averaging, we model the X-ray costs and the rMRI costs with the single parametric models included in $\mathcal{M}$, and with the mixture model of Conigliani and Tancredi (2005a). In fact, the standard parametric analysis is extremely helpfull to illustrate how sensitive inference about cost-effectiveness is

Table 3: Low back pain trial: posterior summaries of mean X-ray cost

Model for costs	$\mathbf{E}(\gamma_1 \mid c_{11}...c_{1n_1})$	$\mathbf{PCI_{0.95}}$	$\pi(M_{l1} \mid c_{11}...c_{1n_1})$
Log-normal	1603	1529; 1678	0
Gamma	1487	1438; 1540	0
Weibull	1548	1447; 1658	0
Loglogistic	1199	1132; 1266	1
GPD	1496	1423; 1579	0
Inverse Gaussian	1329	1281; 1376	0
BMA	**1199**	**1132;1266**	-
Mixture model	**1555**	**1301;1837**	-

Table 4: Low back pain trial: posterior summaries of mean rMRI cost

Model for costs	$\mathbf{E}(\gamma_2 \mid c_{21}...c_{2n_2})$	$\mathbf{PCI_{0.95}}$	$\pi(M_{l2} \mid c_{21}...c_{2n_2})$
Log-normal	2022	1941; 2105	0
Gamma	2159	2079; 2233	0.86
Weibull	2058	1914; 2209	0
Loglogistic	1362	1290; 1445	0.14
GPD	2019	1904; 2148	0
Inverse Gaussian	2104	2036; 2172	0
BMA	**2052**	**1326;2230**	-
Mixture model	**2377**	**1840;3051**	-

to the choice of the model for costs, and how an approach based on Bayesian model averaging can be used to overcome this problem. On the other hand, although in the applications at least one of the distributions in $\mathcal{M}$ will have a positive posterior model probability, there is no guarantee that either of them fit the data well. For this reason we find interesting to compare the results obtained with Bayesian model averaging also with those obtained with the mixture model of Conigliani and Tancredi (2005a), that does not require any assumption about the distribution of costs, and allows inference on the mean cost to take account of the uncertainty about the tail; details of the model and of the prior assumptions can be found in Tancredi *et al.* (2006) and in Conigliani and Tancredi (2005a, 2008).

Consider first the results of the standard parametric analysis. According to the posterior model probabilities shown in Table 3 and Table 4, while in the X-ray group the data definitely support the

Table 5: Low back pain trial: posterior summaries of cost differential and effect differential

Model for costs	$\mathbf{E}(\boldsymbol{\Delta_c}\,\vert D)$	$\mathbf{PCI_{0.95}}$	$\mathbf{P}(\boldsymbol{\Delta_c}>\mathbf{0}\,\vert D)$	$\mathbf{E}(\boldsymbol{\Delta_e}\,\vert D)$	$\mathbf{PCI_{0.95}}$	$\mathbf{P}(\boldsymbol{\Delta_e}>\mathbf{0}\,\vert D)$
BMA	853	116; 1056	1.00	-0.12	-1.6;1.3	0.43
Mixture	822	212; 1538	1.00	0.02	-1.4;1.5	0.51

log-logistic distribution, in the rMRI group plausible models are the gamma (with the highest posterior probability) and the log-logistic. These two models, however, lead to rather different posterior summaries of the mean rMRI cost, and therefore to rather different conclusions in terms of the cost differential Δ_c. In fact, both models suggest that there is evidence of a higher mean cost in the rMRI group, but the strengh of this evidence depends on which model we assume for the data. At one end, if we assume the log-logistic for X-ray costs and the log-logistic for rMRI costs, for the mean cost difference we obtain a point estimate of \$163, and a posterior credible interval that includes the value $\Delta_c = 0$. At the other end, if we assume the log-logistic for X-ray costs and the gamma for rMRI costs, for the mean cost difference we obtain a point estimate of \$960, and a posterior credible interval all on the positive line. It follows that in terms of the analysis of costs and of cost-effectiveness here different models with non-zero posterior probability produce rather different results. And this is exactly a situation where it seems appropriate to apply model averaging: instead of choosing between different models, and then studying the sensitivity of the conclusions in terms of cost-effectiveness, model averaging takes into account all models which are plausible for the data. Notice that in the particular study we are presenting, taking into account both the gamma and the log-logistic distribution for rMRI costs results is a rather wide credible interval for γ_2: with respect to the interval obtained under the gamma model, the one obtained with BMA includes also smaller values of γ_2, that receive a positive posterior probability under the log-logistic model.

Consider now the comparison between the results of Bayesian model averaging and those obtained with the mixture model of Conigliani and Tancredi (2005a). Looking first at Table 3 and Table 4 we notice that the point estimates of γ_1 and γ_2 obtained with the mixture model are higher than those obtained with BMA, and the posterior credible intervals for γ_1 and γ_2 obtained with the mixture model are wider than those obtained with BMA, so that under both treatment groups the mixture model seems to give more weight to the upper tail of the cost distribution than any of the models included in $\mathcal{M}$. Then looking at Table 5 and at the two Cost-Effectiveness Acceptability Curves in Figure 2, we see that also in terms of cost-effectiveness the two approaches lead to different conclusions. In particular, although the point estimates of Δ_c and the probability that $\Delta_c > 0$ are nearly identical (as a consequence of the fact that BMA underestimates the mean costs with respect to the mixture model under both treatments), the

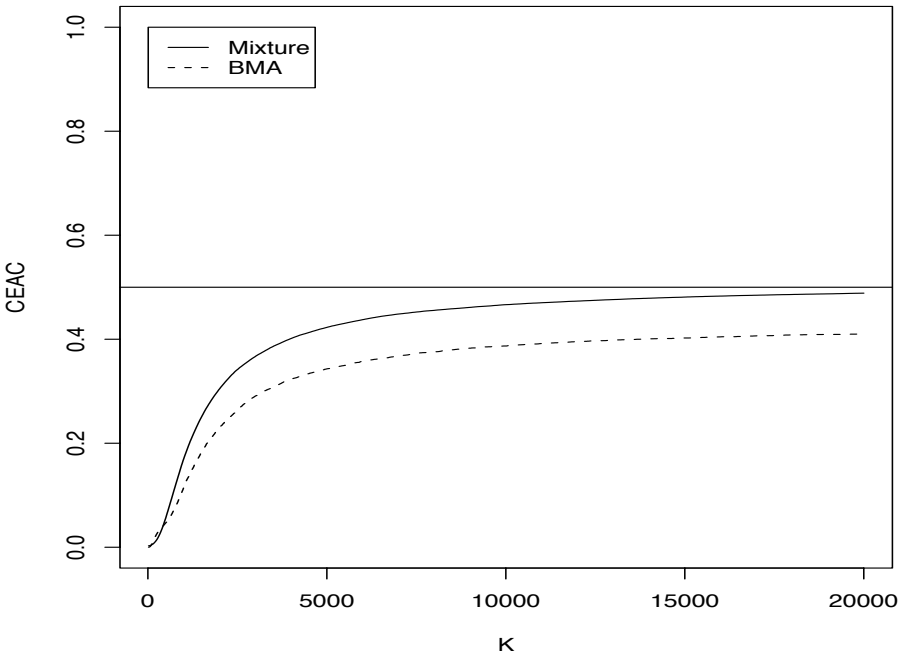

Figure 2: Low back pain trial: cost-effectiveness acceptability curves

probability $Q(K)$ that rMRI is cost-effective is always lower under BMA than under the mixture model. And if we look at the values of K where $Q(K)$ is at least 0.5, we find $K \geq \$30.000$ under the mixture model, while no such values of K exist under BMA. As pointed out in Conigliani and Tancredi (2008), in such cases is very difficult to decide which results one should believe, since these differences are mainly related to the way the two methods deal with model uncertainty. This issue will be addressed further in the final Section.

5 Discussion

Most of the recent literature on cost-effectiveness analyses of clinical trial data agrees that inferences are significally sensitive to the choice of the model for costs, and especially to how the upper tail of the cost distribution beyond the observed data is modelled. The proposal of Conigliani and Tancredi (2005a) to overcome this problem combines the semi-parametric approach to density estimation based on mixture models and the semi-parametric approach to tail estimation based on extreme value theory; the result is a very flexible model able to fit data set with very different shapes both in the bulk of data and in the tail,

but there is a price to pay for so much flexibility in terms of precision and efficiency of the corresponding inferences.

In this paper we have considered an approach based on Bayesian model averaging, that is in the spirit of the sensitivity analyses advocated by Thompson and Nixon (2005). It requires the specification of a set $\mathcal{M}$ of plausible models for cost data, but instead of studying how the conclusions change with the different models, it takes into account the inferences obtained with all the models in $\mathcal{M}$ that have a non-zero posterior probability.

An approach based on Bayesian model averaging for cost-effectiveness analyses in health economics was already proposed in Conigliani and Tancredi (2008), but it involved proper prior distributions for the unknown parameters of the different parametric models. Here we have considered the particular case of weak prior informations, and the main consequence of this assumption is that standard Bayesian model averaging cannot be applied. However in accordance with the theory fractional Bayes factors, we have introduced a new procedure that can deal with improper priors. The computational issues that we encountered were dealt with Path Sampling together with an adaptive Metropolis-Hastings algorithm.

We believe that a BMA type procedure that can be applied even when there is not enough prior knowledge to elicit proper prior distributions for the parameters is particularly relevant especially in a setting like this, where it is difficult to determine plausible statistical models. In this sense, the present approach is closer to the semi-parametric approach of Conigliani and Tancredi (2005a), that does not require any assumption about the distribution of costs. However even if *a priori* we are not introducing any informations, the specification of the set $\mathcal{M}$ has nevertheless the effect of reducing the model uncertainty. It follows that Bayesian model averaging will generally lead to smaller intervals than the semi-parametric approach, but the characteristics of these intervals significantly depends on which models are included in the procedure, so that particular care should be devoted to specifying $\mathcal{M}$.

Acknowledgments I would like to thank Dr. Jerry Jarvik and Dr. Will Hollingworth for permission to use the low back pain trial data in our methodological study. The full costs and effects results from the trial are published elsewhere (Jarvik *et al.*, 2003); my retrospective analysis of the data is for expository purposes only. I am also grateful to Andrea Tancredi and Antonio Parisi for some usefull comments and ideas.

References

Al MJ, van Hout BA. A Bayesian approach to economic analyses of clinical trials: the case of Stenting versus Balloon Angioplasty. *Health Economics* 2000; 9; 599-609.

Berger, J.O., Pericchi, L.R. (1996) The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association*, 91, 109-122.

Carlin, B.P. and Chib, S. (1995). Baysian model choice via Markov chain Monte Carlo. *J. Roy. Statist. Soc. B* 57, 473-484.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *J. Amer. Statist. Assoc.* 90, 1313-1321.

Chib S, Jeliazkov I. Marginal likelihood from the Metropolis-Hastings output, *Journal of the American Statistical Association* 2001; 96; 270-281.

Conigliani C, Tancredi A. Semi-parametric modelling for costs of health care technologies, *Statististics in Medicine* 2005a; 24; 3171-3184.

Conigliani C, Tancredi A. A Bayesian semi-parametric approach for cost-effectiveness analysis in health economics, Working Paper del Dipartimento di Economia, Università Roma Tre, 2005b; 46.

Conigliani C, Tancredi A. A Bayesian model averaging approach for cost-effectiveness analyses. *Health Economics*, to appear, 2008.

De Bruijn, N.G. (1961). *Asymptotic Methods in Analysis*, Amsterdam, North-Holland.

Dinh P, Zhou X. Nonparametric Statistical Methods for Cost-Effectiveness Analyses, *Biometrics* 2006; 62; 576-588.

Fryback DG, Chinnis JO, Ulvila JW. Bayesian cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care* 2001; 17; 83-97.

Gelfand, A.E. and Dey, D.K. (1994). Bayesian model choice: asymptotics and exact calculations, *J. Roy. Statist. Soc. B* 56, 99-118.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities, *J. Amer. Statist. Assoc.* 85, 398-409.

Gelman, A. and Meng, X-L. (1994). Path sampling for computing normalizing constants: identities and theory. Technical Report 376, Department of Statistics, University of Chicago.

Gelman, A. and Meng, X-L. (1998). Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling. *Statist. Sci.* 13, 163-185.

Genz, A. and Kass, R.E. (1993). Subregion Adaptive Integration of Functions Having a Dominant Peak. Technical Report, Carnegie Mellon University, Dep. of Statistics.

Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica* 57, 1317-1339.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711-732.

Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial (with discussion), *Statistical Science* 1999; 14; 382-417.

Jarvik JG, Hollingworth W, Martin B, et al. Rapid magnetic resonance imaging vs. radiographs for patients with low back pain. *Journal of the American Medical Association* 2003; 289; 2810-18.

Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* 90, 773-795.

Meng, X.L. and Wong, W.H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* 6.

Newton, M.A. and Raftery, A.E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap, *J. Roy. Statist. Soc.* B 56, 3-48.

Nixon RM, Thompson SG. Parametric modelling of cost data in medical studies. *Statististics in Medicine* 2004; 23; 1311-1331.

O'Hagan, A. (1991). Bayes-Hermite quadrature. *J. Statist. Planning and Inference* 29, 245-260.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). *J. Roy. Statist. Soc. B* 57, 99-118.

O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test* 6, 101-118.

O'Hagan A, Forster J. Bayesian Inference, Kendall's Advanced Theory of Statistics, II edn, vol. 2B, Arnold, London; 2004.

O'Hagan A, Stevens JW. A framework for cost-effectiveness analysis from clinical trial data. *Health Economics* 2001; 10; 302-315.

O'Hagan A, Stevens JW. Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies. *Statistical Methods in Medical Research* 2002; 11; 469-490.

O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Economics* 2003; 12; 33-49.

O'Hagan A, Stevens JW, Montmartin J. Inference for the Cost-Effectiveness Acceptability Curve and Cost-Effectiveness Ratio, *Pharmacoeconomics* 2000; 17; 339-349.

Raftery, A.E. (1996). Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models. *Biometrika* 83, 251-266.

Rosenkrantz, S. (1992). The Bayes Factor for Model Evaluation in a Hierarchical Poisson Model for Area Counts. Ph.D. Thesis, University of Washington, Dep. of Biostatistics

Smith, A.F.M. (1991). Bayesian computational methods, *Phil. Tr. R. Soc. Lond. A*, 337, 369-386.

Smith, A.F.M., Skene, A.M., Shaw, J.E.H., Naylor, J.C. and Dransfield, M. (1985). The implementation of the Bayesian paradigm, *Commun. Statist. - Theor. Meth.*, 14(5), 1079-1102.

Stinnett AA, Mulahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis, *Medical Decision Making* 1998; 18; 68-80.

Tancredi A, Anderson CW, O'Hagan A. Accounting for threshold uncertainty in extreme value estimation. *Extremes* 2006; 9; 87-106.

Tanner, M.A. (1993). *Tools for Statisical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, II Ed., New York, Springer-Verlag.

Tanner, M.A. and Wong, W. (1987). The Calculation of Posterior Distributions by Data Augmentation (with discussion), *J. Amer. Statist. Assoc.*, 82, 528-550.

Thompson SG, Nixon RM. How sensitive are cost-effectiveness analyses to choice of parametric distributions? *Medical Decision Making* 2005; 25; 416-423.

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities, *J. Amer. Statist. Assoc.,* 81, 82-86.

van Hout BA, Al MJ, Gordon GS, et al. Costs, effects and C/E ratios alonside a clinical trial. *Health Economics* 1994; 3; 309-19.