

A Bayesian approach to Markov modelling in cost-effectiveness analyses: application to taxane use in advanced breast cancer

Nicola J. Cooper, Keith R. Abrams, Alex J. Sutton, David Turner and Paul C. Lambert

University of Leicester, UK

[Received February 2002. Revised November 2002]

Summary. The paper demonstrates how cost-effectiveness decision analysis may be implemented from a Bayesian perspective, using Markov chain Monte Carlo simulation methods for both the synthesis of relevant evidence input into the model and the evaluation of the model itself. The desirable aspects of a Bayesian approach for this type of analysis include the incorporation of full parameter uncertainty, the ability to perform all the analysis, including each meta-analysis, in a single coherent model and the incorporation of expert opinion either directly or regarding the relative credibility of different data sources. The method is described, and its ease of implementation demonstrated, through a practical example to evaluate the cost-effectiveness of using taxanes for the second-line treatment of advanced breast cancer compared with conventional treatment. For completeness, the results from the Markov chain Monte Carlo simulation model are compared and contrasted with those from a classical Monte Carlo simulation model.

Keywords: Bayesian methods; Breast cancer; Cost-effectiveness decision analysis; Elicitation; Markov models; Meta-analysis; Taxanes

1. Introduction

Decision analytical modelling is widely used by health economists to estimate the cost-effectiveness of health care interventions with the aim of providing information to allow scarce health care resources to be allocated efficiently (Briggs and Sculpher, 1998). These types of model have a range of uses, which include the synthesis of data from a variety of sources to produce the cost-effectiveness results of interest. For example, when economic evaluations are conducted alongside trials, modelling techniques are often used to extrapolate the primary data beyond the end point of the trial (Brennan and Akehurst, 2000; Briggs, 2000).

Decision models are used in economic evaluation studies to evaluate the complex process that is often associated with the implementation of health care interventions (Lilford and Royston, 1998). In particular, Markov decision models provide a technique for analysing events that are repeatable (e.g. mental health treatment) or events that play out over an extended period of time (e.g. the progression of cancer) (Keeler, 1996; Sonnenberg and Beck, 1993). Unlike other applications of Markov modelling in medical research, economists are interested in both the resource use and health outcome of health care interventions. Markov models provide a way of handling both costs and outcomes simultaneously in a simple and intuitive manner. Such models are often deterministic, using point estimates with no associated measure of uncertainty.

Address for correspondence: Nicola J. Cooper, Department of Epidemiology and Public Health, University of Leicester, 22–28 Princess Road West, Leicester, LE1 6TP, UK.
E-mail: njc21@le.ac.uk

However, more sophisticated probabilistic (i.e. consisting of empirical estimates for parameters of interest with corresponding uncertainty) models can be developed (Briggs, 2000).

To date Markov models in economic evaluation have almost exclusively been analysed by using classical statistical approaches; one of the few exceptions has been Parmigiani *et al.* (1997). The aim of this paper is to demonstrate how probabilistic Markov models can be implemented from a Bayesian perspective, using Markov chain Monte Carlo (MCMC) simulation for both the synthesis of relevant evidence input into the model and the evaluation of the model itself. The relative merits of the Bayesian approach over standard classical methodology, including the ability to perform each meta-analysis of the relevant parameters in a single coherent model, are presented and demonstrated via a detailed application to investigate the cost-effectiveness of using taxanes for the second-line treatment of advanced breast cancer compared with conventional treatment.

The paper is organized with Section 2 considering the background to the treatment of advanced breast cancer, whereas Section 3 introduces the Markov model that is developed to assess the cost-effectiveness of using taxanes compared with conventional treatment. The model is then analysed by using a Bayesian approach and for completeness the results are compared and contrasted with a standard classical approach in Section 4. Section 5 discusses some of the issues that have been raised and outlines specific areas for further research. More details about the structure of the model, prior information and parameter sources and estimates can be found at <http://www.prw.le.ac.uk/epidemiology/personal/njc21/index.html>.

2. Use of taxanes in the second-line treatment of advanced breast cancer—background

In the UK the incidence and mortality of breast cancer are the highest in the world. There are approximately 35 000 new cases and over 13 000 breast-cancer-related deaths per annum (Cancer Research Campaign, 1996). Tumour cells are distributed through the body via the blood and lymphatic systems and may develop into secondary tumours or metastases. For those women who develop metastases the prognosis is poor and metastatic disease is often considered incurable (Scottish Health Purchasing Information Centre, 1997). For such people, treatment only provides a temporary control of the growth of the cancer. The main goals of the treatment are

- (a) to relieve symptoms with as few side-effects as possible and
- (b) to extend the duration of high quality life (Fornier *et al.*, 1999).

The current treatment options for metastatic breast cancer include endocrine therapy, anthracyclines (e.g. doxorubicin and epirubicin), cyclophosphamide, methotrexate, fluorouracil, mitomycin, mitoxantrone and the taxanes. Taxanes are a class of anti-cancer (chemotherapy) drugs, originally derived from the bark of the Pacific yew, *Taxus brevifolia*. There are two types of taxane—paclitaxel (Taxol®) and docetaxel (Taxotere®)—both of which have similar mechanisms of action. At present both taxanes are used in the treatment of ovarian and breast cancer and can be used as either the first-line treatment (initial systematic therapy after surgery) or the second-line treatment (if disease persists or relapses) (Lister-Sharp *et al.*, 2000). In this illustrative example we focus on the cost-effectiveness of docetaxel as a second-line treatment for metastatic breast cancer compared with conventional treatment (assumed here to be doxorubicin). To date these two treatments have only been compared directly in one published clinical trial (Chan *et al.*, 1999).

3. Markov model

3.1. Model description

To investigate the cost-effectiveness of using taxane chemotherapy treatment (i.e. docetaxel) in place of conventional chemotherapy treatment (i.e. doxorubicin) for advanced breast cancer, a four-stage probabilistic Markov model was developed based on an earlier model by Brown and Hutton (1998). A Markov model assumes that individuals are always in one of a finite number of states of health referred to as Markov states and that health changes from state to state according to a set of transition probabilities (Sonnenberg and Beck, 1993). These transition probabilities depend only on the current health state that the person is in and not on their previous health states (the Markov assumption) (Keeler, 1996). This particular model consisted of four health states:

- (a) *response*—complete and partial (reduced by at least 50% in size) tumour disappearance;
- (b) *stable*—no change;
- (c) *progressive*—tumour growth or spread to other sites;
- (d) *death*—the absorbing state (i.e., once entered, it is never left).

The model was developed so that one cycle of the model represented 3 weeks to coincide with the intervals of chemotherapy treatment. Note that other cycle lengths could have been chosen; however, evaluating the model in terms of shorter cycles or even continuous time would be extremely computer intensive. It was assumed that individuals were only allowed a maximum of seven cycles of treatment owing to problems with cumulative toxicities that are associated with the chemotherapy agents (stages 1, 2 and 3). Individuals were assumed to discontinue treatment before seven cycles if either they moved into the 'progressive' health state or suffered major toxic reactions from the treatment. The model followed individuals for a maximum of 28 further cycles after the end of the treatment period (stage 4) as by this time the majority of individuals had reached the absorbing state (i.e. death). Note that individuals could only move from the 'stable' health state to the 'response' health state during the treatment cycles and thereafter only digress to the 'progressive' health state and 'death'.

Fig. 1 summarizes all the possible health state pathways at the end of each cycle for stages 1–4 (a more detailed diagram can be found at <http://www.prw.le.ac.uk/epidemiology/personal/njc21/index.html>). The main assumptions of the model were that

- (a) individuals were not classified into the various health states until stage 2 of the model (i.e. cycle 3—the third round of treatment),
- (b) once the progressive health state had been entered individuals either remained in this state or moved to the death state,
- (c) individuals could only enter the death state from the progressive health state (except in cycles 1 and 2) and
- (d) minor toxicities were treatable and chemotherapy continued.

To avoid overcomplicating the model, the possibility of reductions of dose and delays in treatment due to toxicities were excluded.

3.2. Model parameters

3.2.1. Transition probabilities

Transition probabilities were obtained from the latest published clinical trial data. Where data were available from more than one source, meta-analysis techniques were used to synthesize

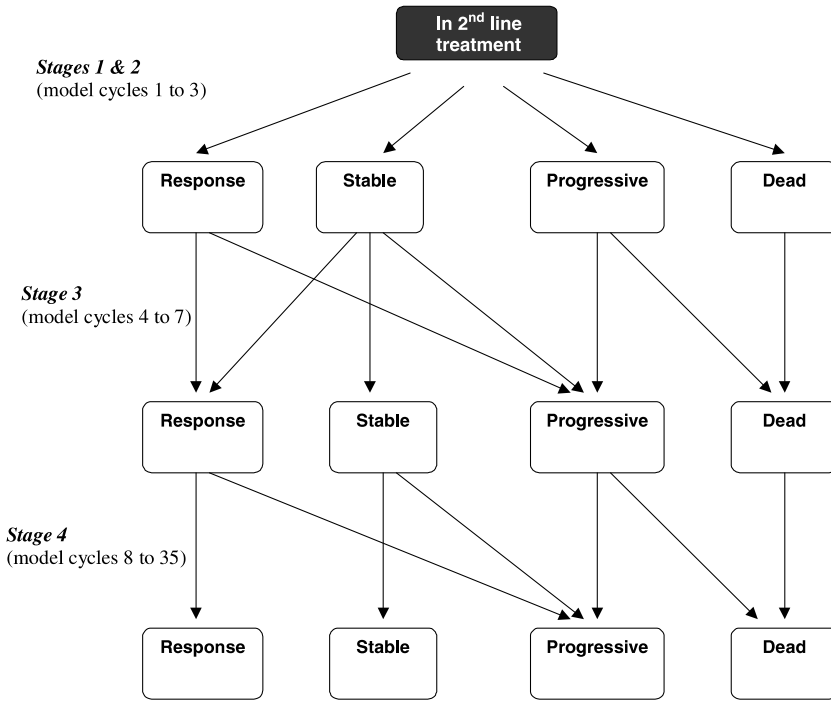


Fig. 1. Simplified diagram of pathways through the model

the evidence (Sutton *et al.*, 2000a). The methods that were used to obtain (pooled) estimates of the transition probabilities and their corresponding uncertainty (in the form of a probability distribution) from the clinical trial data are described below.

Random-effect meta-analysis models were used to combine the data throughout. The generic form of this two-level hierarchical model can be expressed as

$$\begin{aligned}
 y_i &\sim \text{normal}(\theta_i, \sigma_i^2) & i = 1, \dots, m, \\
 \theta_i &\sim \text{normal}(\mu, \tau^2),
 \end{aligned}
 \tag{1}$$

where y_i is the effect size estimate of interest from the i th of m studies being combined, θ_i is the true underlying effect size for the i th study with corresponding variance σ_i^2 , μ is the overall pooled effect size estimate and τ^2 is the estimate of between-study heterogeneity. In a classical approach, τ^2 is often estimated by using a ‘method-of-moments’ approach from the data, whereas σ_i^2 is frequently assumed to be known and estimated by using the within-study observed sample variance (DerSimonian and Laird, 1986). However, the estimation of μ typically ignores the uncertainty that is associated with the estimate of τ^2 , though profile likelihood methods exist to overcome this limitation (Hardy and Thompson, 1996). For the Bayesian analysis, σ_i^2 is also assumed to be known; however, the uncertainty in estimating τ^2 from the data is explicitly taken into account by specifying it as a random variable. Hence prior distributions are required for μ and τ^2 in the Bayesian analysis; these are discussed in Section 3.4.

When the outcome of interest was based on binary data, required for estimating the probability of

- (a) responding to treatment,
- (b) moving to the progressive health state at the end of stage 2 and
- (c) different adverse events,

meta-analyses were performed using the log-odds scale, as this should provide a measure which is approximately normally distributed (Sutton *et al.*, 2000a). When outcomes were measured on a continuous scale, required for estimating median times to different events, the weighted (by the inverse of the variance) mean of the medians outcome was used for the analysis (Sutton *et al.*, 2000a). Separate meta-analyses were carried out for doxorubicin and docetaxel (the two treatment options) and normal distributions were assumed throughout. The specific outcome measures that were used are listed in Tables 1 and 2.

The pooled estimates that were obtained from the meta-analyses were then converted into transition probabilities of model events. Where the available information was expressed in terms of a median time to an event, transition probabilities were estimated via rates (Miller and Homan, 1994):

$$R = \frac{-\ln\{1 - P(t_0, t_j)\}}{j} \tag{2}$$

Here $P(t_0, t_j)$ is the cumulative probability between times t_0 and t_j which is estimated with uncertainty via $P(t_0, t_j) = -\ln(0.5)/\mu$ where μ is estimated from the relevant meta-analysis by using equation (1), R is the rate per cycle and j represents the relevant number of equal time intervals required by the model (for example if $P(t_0, t_j)$ represented the 12-month cumulative probability then to obtain the rate per month j would equal 12). Then the transition probability per cycle, P , is given by

$$P = 1 - \exp(-R) \tag{3}$$

This assumes that the instantaneous transition probability remains constant during the entire period (i.e. exponential reduction in the eligible population).

Where the information was expressed as a proportion (e.g. the proportion responding or the proportion progressing), transitional probabilities were estimated by (Miller and Homan, 1994)

$$P = 1 - \{1 - P(t_0, t_j)\}^{1/j} \tag{4}$$

Table 1. Pooled results from Bayesian meta-analyses: times

	Times in weeks—means (95% credible intervals) across studies for the following treatments:			
	Number of studies	Docetaxel	Number of studies	Doxorubicin
Median progression-free time	3†‡§	25 (15–33)	2†§§	22 (15–29)
Median time to response from stable	1†	12 (6–18)	1†	23 (17–29)
Median time to progressive from response	1*	35 (29–41)	1§§	40 (34–46)
Median overall survival time	3†‡§	53 (35–74)	2†§§	56 (37–99)

†Chan *et al.* (1999).
 ‡Nabholtz *et al.* (1999).
 §Sjöström *et al.* (1999).
 §§Bontenbal *et al.* (1998).
 *Bonnetterre *et al.* (1997).

Table 2. Pooled results from Bayesian meta-analyses: probabilities

	<i>Probabilities—means (95% credible interval) across studies for the following treatments:</i>			
	<i>Number of studies</i>	<i>Docetaxel</i>	<i>Number of studies</i>	<i>Doxorubicin</i>
Response rate	4†‡§§§§	0.43 (0.29–0.58)	2†*	0.34 (0.17–0.58)
Moving directly to progressive at stage 2	1†	0.13 (0.08–0.18)	1†	0.22 (0.16–0.29)
Infections or febrile neutropenia	3†‡§	0.18 (0.04–0.56)	1†	0.16 (0.11–0.22)
Hospitalized following infection or febrile neutropenia	1**	0.08 (0.05–0.11)	0	0.08 (0.05–0.11)
Dying from infections or febrile neutropenia	1†	0.01 (0.00–0.02)	2†*	0.04 (0.00–0.46)
Discontinue treatment due to adverse event	3†‡§	0.16 (0.03–0.49)	2†*	0.18 (0.05–0.40)
Neutropenia grades 3 and 4	2†‡	0.94 (0.82–0.98)	1†	0.89 (0.84–0.93)
Anaemia grades 3 and 4	2†§	0.03 (0.00–0.28)	1†	0.16 (0.11–0.22)
Diarrhoea grades 3 and 4	3†‡§	0.09 (0.06–0.14)	2†*	0.02 (0.00–0.32)
Stomatitis grades 3 and 4	3†‡§	0.08 (0.04–0.14)	1†	0.17 (0.11–0.25)
Vomiting grades 3 and 4	2†‡	0.03 (0.00–0.12)	2†*	0.10 (0.02–0.35)
Fluid retention grades 3 and 4	3†‡§	0.05 (0.02–0.12)	1†	0.00 (0.00–0.02)
Cardiac toxicity grades 3 and 4	1†	0.00 (0.00–0.02)	2†*	0.05 (0.01–0.24)

†Chan *et al.* (1999).
 ‡Nabholtz *et al.* (1999).
 §Sjöström *et al.* (1999).
 §§Bonnetterre *et al.* (1997).
 *Bontenbal *et al.* (1998).
 **O'Brien *et al.* (1999).

where $P(t_0, t_j)$ and j are defined as above in equation (2). Again it is assumed that the true transition probability remains constant over the time period.

Note that the estimated transition probabilities all had uncertainties associated with them and hence were represented by probability distributions. These distributions were defined by the propagated parameter uncertainty relating to the meta-analyses results that they were based on. In the case of the Bayesian analysis, the exact posterior distribution for the pooled parameters from the meta-analysis were used, whereas, for a classical analysis, the log-odds of an event and times to events would be assumed to be normally distributed.

3.2.2. Utilities

Utilities are a measure of preference between health states and were obtained for this model by pooling information from published cost–utility studies of taxanes. The utility data for the model were extracted directly from published studies where the quality-of-life evaluation had been conducted on oncology doctors and nurses, rather than on the cancer patients themselves, using either health state descriptions (Feeny *et al.*, 1995) or standard gamble methodology (Furlong *et al.*, 1990). Where necessary, the authors of the published studies had re-expressed their data on a utility scale from 0 to 1 (where 0 represents death and 1 represents full health) by using weighting factors. The published sources for these data did not report the degree of variability between individuals for these utility values, only the mean values. Hence, to express uncertainty and variability in these estimates, the standard error of the between-study means were calculated. No data were available for the utility associated with progressive disease with

toxicity so it was necessary to extrapolate from the utilities for the other health states to estimate this quantity. These utility estimates were assumed to have beta distributions, the parameters of which were obtained by using methods-of-moments estimation (Briggs, 2000). Utilities were measured in quality-adjusted life years (QALYs).

3.2.3. Costs

The resource uses that were associated with the two treatment regimes were obtained from a combination of published information, where available, and clinical opinion otherwise. Drug dosages were obtained from Lister-Sharp *et al.* (2000). Unit cost information was obtained from a range of sources (see <http://www.prw.le.ac.uk/epidemo/personal/njc21/index.html> for more details). Hospital costs (i.e. febrile neutropenia or infections, severe fluid retention or cardiac toxicity), the main contributor of health care costs after treatment, were assigned normal distributions informed by the range of costs provided by the Department of Health reference costs (Department of Health, 1999). All other cost information was entered as deterministic because no measures of uncertainty were available; however, the sensitivity of the results to certain cost estimates assigned were investigated (see Sections 3.5 and 4.3). Note that, although some of these data were derived from opinions, they were treated as data in the model and do not contribute to any prior distributions (Section 3.4). All costs were measured in 1999 UK pounds sterling.

3.3. Evaluating the Markov model

In both the classical and the Bayesian approaches, the type of analysis that was undertaken was a ‘cohort analysis’ (Sonnenberg and Beck, 1993). The model was set up to simulate the prognosis of a hypothetical cohort of 1000 individuals on doxorubicin treatment and 1000 individuals on docetaxel treatment (the sensitivity of the results to the number of individuals run through the model was investigated as part of the sensitivity analysis (see Sections 3.5 and 4.3)). For each cycle of the model, individuals moved between health states according to the associated transition probabilities. This resulted in a new allocation of the cohort between the various health states for the subsequent cycle. The model assumed that all individuals initially entered the model at the initiation of treatment (initially not allocated to a health state).

Since the model that was implemented was fully probabilistic, different values for the underlying parameters were sampled from specified ranges (informed by applying plausible ranges of values) or distributions (informed by the meta-analyses—see Section 3.2). New parameter values were sampled for each new iteration of the simulated model (i.e. one iteration consisted of an evaluation of the full 35 model cycles). In the classical model, constructed for comparative purposes, Monte Carlo simulation (Briggs, 2000; Sonnenberg and Beck, 1993) was used to sample from the specified ranges or distributions, whereas for the Bayesian model MCMC simulation (specifically Gibbs sampling) (Gilks *et al.*, 1996) was adopted. The model was evaluated by averaging parameter values of interest (see below) over many iterations, allowing uncertainty in model parameters to be accounted for.

For each iteration of the simulated model, the cost and utility accrued for each cycle, referred to as the *cycle sums*, were calculated for each treatment regime separately by the formulae (Sonnenberg and Beck, 1993)

$$\text{cycle sum (utility)} = \sum_{s=1}^N n_s U_{sk}, \quad (5)$$

$$\text{cycle sum (cost)} = \sum_{s=1}^N n_s C_{sk} \quad (6)$$

where N is the number of health states, n_s is the number of individuals in state s (where $\sum_{s=1}^N n_s = 1000$) and k represents the two treatment groups (docetaxel and doxorubicin). U_s is the cycle utility of health state s (i.e. the utility that is associated with spending one cycle in a particular health state) and C_s is the cycle cost of health state s (i.e. the cost that is associated with spending one cycle in a particular health state). At the end of each iteration, the *cumulative utility* and *cumulative cost* were obtained by adding the cycle sums together. The mean costs \bar{C}_k and utilities \bar{U}_k for each iteration were then calculated by dividing the cumulative utility and cumulative cost by 1000 individuals. The mean incremental cost

$$\Delta C = \bar{C}_{\text{DOC}} - \bar{C}_{\text{DOX}} \quad (7)$$

and mean incremental effectiveness

$$\Delta U = \bar{U}_{\text{DOC}} - \bar{U}_{\text{DOX}} \quad (8)$$

of docetaxel over doxorubicin were then evaluated at every iteration. Finally, the Monte Carlo expectation across all iterations of the mean incremental cost and effectiveness (i.e. the mean of ΔC and ΔU) denoted by $\Delta \bar{C}$ and $\Delta \bar{U}$ was calculated. This allows the incremental net monetary benefit (NMB) to be calculated as outlined by Stinnett and Mullahy (1998):

$$\text{incremental NMB} = R_c \Delta \bar{U} - \Delta \bar{C} \quad (9)$$

where $\Delta \bar{U}$ and $\Delta \bar{C}$ are the mean incremental utility and mean incremental cost respectively, and R_c is the ceiling ratio (i.e. the amount that decision makers are willing to pay per additional QALY gained). If the incremental NMB is positive then it suggests that the new intervention represents good 'value for money' given the chosen ceiling ratio, whereas a negative value suggests that the intervention is cost ineffective. The final results can then be expressed on a cost-effectiveness acceptability curve with the probability cost effective (the number of iterations for which the incremental NMB is positive divided by the total number of iterations) calculated for a range of values for R_c (Stinnett and Mullahy, 1998).

3.4. Prior distributions (for Bayesian analysis)

3.4.1. 'Vague prior distributions' model

In adopting a Bayesian approach it is necessary to specify prior distributions for all unknown parameters in the simulation model, which allows the incorporation of information that is external to that being analysed directly (Carlin and Louis, 2000). However, prior distributions can be specified as vague or non-informative relative to the observed data if no external evidence is available.

In this model, vague normal prior distributions were placed on all the population effect parameters (derived from the meta-analyses as outlined in equation (1) and denoted by μ). All these prior distributions were centred at zero and had a large standard deviation of 1000, making them approximately flat across the range of feasible values. The between-study variance parameters (denoted by τ^2 in equation (1)) in each meta-analysis were given inverse gamma distributions with both scale and shape parameters set to be 0.001, with the intention of being minimally informative, though as with any Bayesian analysis the robustness of the results that are obtained to such prior distributional assumptions should be investigated via a sensitivity analysis (Spiegelhalter *et al.*, 2000). See Natarajan and Kass (2000) for further discussion regarding the choice of vague prior distributions.

3.4.2. 'Informative prior distributions' model

The above analysis used vague prior distributions because of the small amount of data-based information that is currently available regarding the effectiveness of docetaxel on patients treated for advanced breast cancer. However, a previous study (Hutton *et al.*, 1996) had elicited the beliefs of 20 oncologists worldwide regarding the 'response rate' and 'time to progressive from response' for both standard and docetaxel treatment by using a 'trial roulette' elicitation approach (Gore, 1987) which has been used in other cancer settings (Abrams *et al.*, 1994; Parmar *et al.*, 1994, 2001). Each oncologist was individually and independently asked to provide a histogram which represented *their* beliefs regarding both the response rate and time to progressive from response for standard and docetaxel treatment separately. To utilize this information in our model, their individual beliefs were then combined by using a linear opinion pooling method (Genest and Zidek, 1986) to produce an overall histogram for both the response rates and time to progressive from response (more details are available at <http://www.prw.le.ac.uk/epidemiology/personal/njc21/index.html>).

Before the response rate prior distributions, for doxorubicin and docetaxel treatment, could be integrated into the model a logit transformation was required, since it was the logit of the response rate that was being modelled. The logit (response rate) was then assumed to follow a normal distribution, which seemed reasonable on the basis of a visual inspection of the data, with mean and variance estimated by the sample statistics; therefore the distributions had means -0.95 and -0.56 , and standard deviations 0.92 and 0.87 for doxorubicin and docetaxel treatment respectively (more details are available at <http://www.prw.le.ac.uk/epidemiology/personal/njc21/index.html>). When converted back onto the proportion scale this translates to response rates of approximately 31% (standard deviation 14.8) and 38% (standard deviation 17.3) for doxorubicin and docetaxel treatment respectively. Hence, on average the oncologists believed that the response rate would be approximately 7% greater for docetaxel compared with doxorubicin; however, the large standard deviations for both response rates indicate considerable uncertainty in these prior beliefs. For expert beliefs regarding the time to progressive from response, normal distributions were assumed with means 28.7 weeks and 29.4 weeks, and standard deviations 14.1 and 15.4 for doxorubicin and docetaxel treatment respectively (more details are available at <http://www.prw.le.ac.uk/epidemiology/personal/njc21/index.html>). Hence, the oncologists believed that the times to progressive from response would be similar for the two treatments; however, the relatively large standard deviations suggest that there is considerable uncertainty in their beliefs. For further discussion on the pooling of distributions obtained from multiple experts see Genest and Zidek (1986).

3.5. Sensitivity analysis

In a complex decision model such as that described above, very many assumptions must be made in terms of both model structure and model inputs. It is therefore important to undertake an extensive sensitivity investigation as a major part of the cost-effectiveness assessment. In this example, one of the main assumptions of the model was the decision to treat some of the cost components as deterministic as no measures of variation were available. The main cost in this category was that of treating neutropenia which had a high incidence in individuals who were undergoing treatment. As part of the sensitivity analysis this cost was varied from £500 to £1500. The sensitivity of the results to the number of individuals in the hypothetical cohort was also investigated by varying the sample size from 500 up to 2000. Finally, to analyse the effect of a reduction in cost of docetaxel in the future if a generic version of the treatment were to be introduced, 50% lower treatment costs for docetaxel were compared with no change in the cost of the standard treatment.

Others assumptions could have also been investigated such as the quality of the studies and the complexity of the model, although such assessments are non-trivial. To confirm that the intended 'vague' prior distributions are not exerting undue influence on the overall results, the model should also ideally be evaluated for a range of prior distributions (Cooper *et al.*, 2002).

4. Model results

For the Bayesian models all the analysis was implemented in one coherent model by using WinBUGS version 1.3. The classical analysis, included below for comparative purposes, was implemented by using the statistical software Intercooled Stata 7.0 for the meta-analyses and Microsoft® EXCEL 2000 for the Monte Carlo simulation.

For the Bayesian analyses, following preliminary test runs, it was decided to use an initial run of 2000 iterations as a 'burn-in' (Gilks *et al.*, 1996), to reach convergence (these values were discarded). The model was run twice: once with vague prior distributions and once using the informative prior distributions as described above. Model convergence is not an issue when using Monte Carlo sampling on which the classical model is based. Inferences for all models were based on 4000 sample iterations.

4.1. Meta-analyses

Results from the meta-analyses (using vague prior distributions) are presented in Table 2 for docetaxel and doxorubicin in terms of probabilities and weighted means. As described in Section 3.2 pooled estimates from the meta-analyses (together with their corresponding uncertainty) were converted into transition probabilities before they could be used in the model.

4.2. Simulation models

Fig. 2 shows the dispersion of the hypothetical cohort of 1000 individuals across the various health states at each time point (weeks) in the model for each treatment group. These results are from the classical model, but the results from the Bayesian models were very similar. By comparing Figs 2(a) and 2(b) it can be observed that, whereas the overall survival time for the two groups is very similar, the time to progression (i.e. the progression-free survival time) is shorter for the doxorubicin group owing to a shorter duration of time spent in the response state.

The overall results from the 4000 iterations for the two approaches are displayed in Fig. 3 in terms of the mean incremental costs ΔC plotted against the mean incremental utilities ΔU . By comparing Figs 3(a) and 3(b) it can be observed that the estimates that were obtained from the Bayesian model (using vague prior distributions) are more dispersed than those of the classical model, indicating the greater parameter uncertainty that has been incorporated into the Bayesian model. The main observation from both plots is that docetaxel costs significantly more than doxorubicin (i.e. nearly all points are above the x -axis) but there is little evidence that docetaxel is more effective than doxorubicin (i.e. the points span fairly evenly either side of the y -axis). Table 3 summarizes this information by showing the overall mean incremental costs $\Delta \bar{C}$ and incremental utilities $\Delta \bar{U}$ for the different methods calculated across all 4000 iterations. It can be observed from Table 3 that, as would be expected, the 95% credible intervals for the Bayesian estimates are wider than the 95% confidence intervals for the classical estimates. Note that there are substantial differences between the point estimates of the two Bayesian models and the classical model.

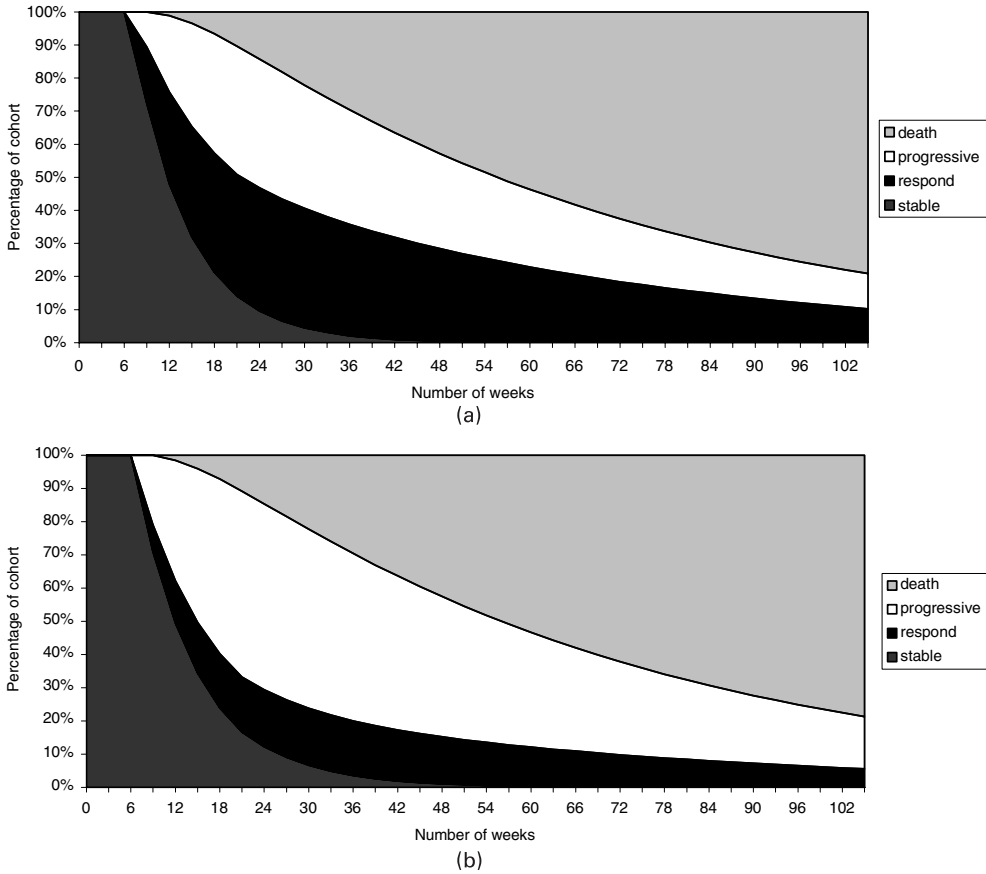


Fig. 2. Expected dispersion of the hypothetical cohort of 1000 individuals over time averaged over iterations (Monte Carlo model): (a) docetaxel treatment group; (b) doxorubicin treatment group

Table 3. Mean incremental costs and utilities

<i>Model</i>	<i>Mean incremental cost</i> $\Delta\bar{C}$ (95% interval) (£)	<i>Mean incremental utility</i> $\Delta\bar{U}$ (95% interval) (£)
Classical (Monte Carlo)	5250 (3175–7262)†	0.047 (–0.110–0.194)†
Bayesian (MCMC)	4468 (1317–7492)‡	0.040 (–0.198–0.270)‡
Bayesian (MCMC) with informative prior distributions§	4438 (1520–7336)‡	0.036 (–0.201–0.251)‡

†95% confidence interval.

‡95% credible interval.

§Informative prior distributions for response rate and time to progressive from response.

Fig. 4 shows the cost-effectiveness acceptability curves for both the Bayesian and the classical approaches. The graph shows that for the R_c -value of, say, £100000 per additional QALY gained the probability that taxanes are more cost effective than the standard treatment is 0.49 for the Bayesian ‘informative prior distribution analysis’ and 0.48 for the classical analysis. The classical estimates are often incorrectly given a direct probabilistic interpretation (for example

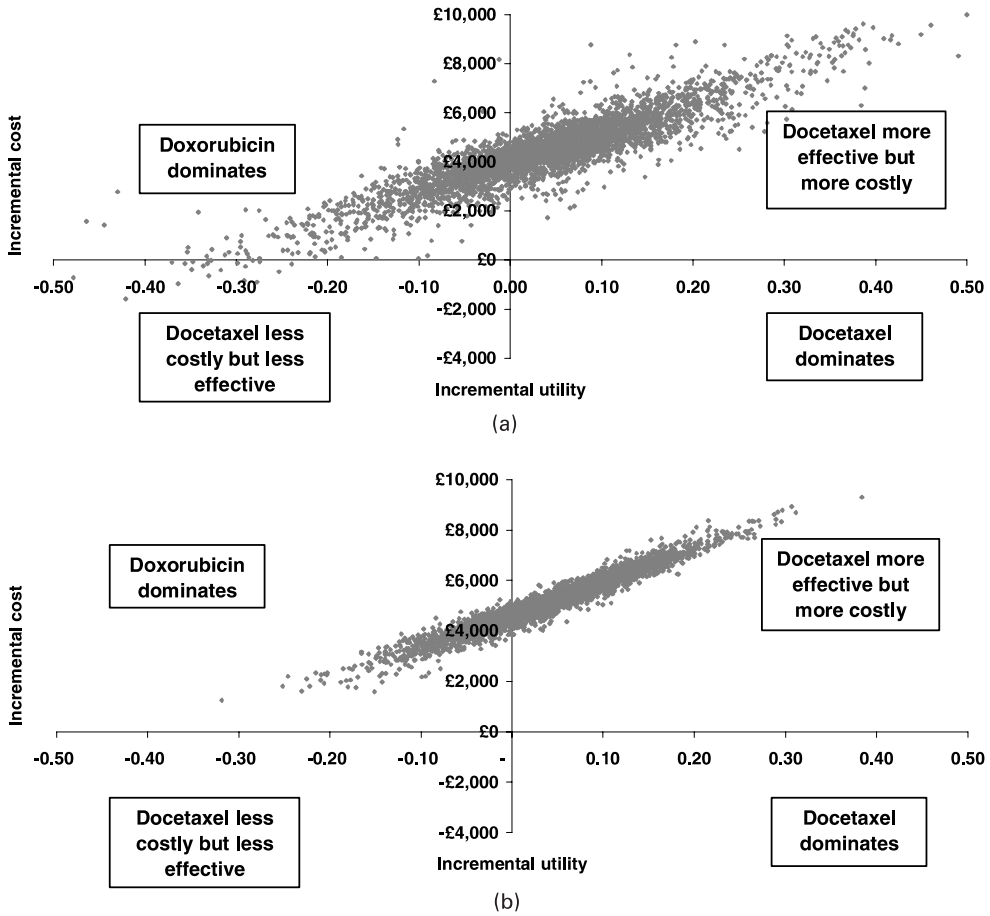


Fig. 3. Model results plotted on the cost-effectiveness plane: (a) Bayesian (MCMC) simulation model; (b) classical (Monte Carlo) simulation model

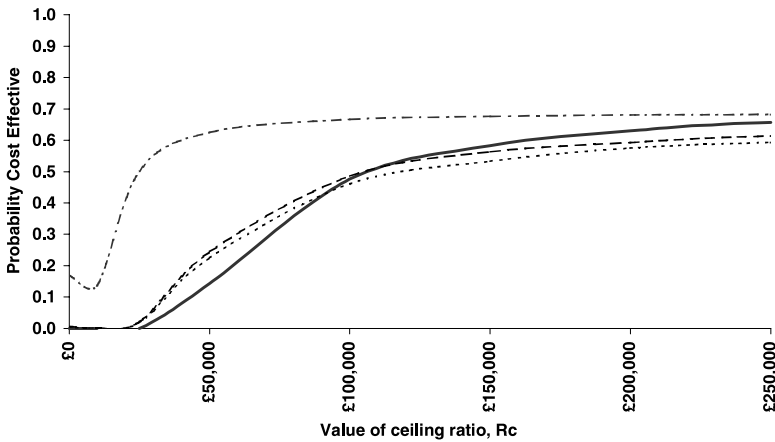


Fig. 4. Cost-effectiveness acceptability curves: —, classical (Monte Carlo) model; ---, Bayesian (MCMC) model; - · - ·, Bayesian (MCMC) with informative prior model; · · · ·, 0.5 docetaxel costs

the value 0.48 is interpreted as that there is a 48% probability that taxanes are cost effective compared with conventional treatment) but such an interpretation is possible from a Bayesian perspective. This direct probabilistic interpretation is another advantage of the Bayesian approach.

4.3. Sensitivity analysis

The results were found to be robust to the number of individuals assumed in the hypothetical cohort and the cost of treating neutropenia. However, when the cost of docetaxel was halved the probability that docetaxel is cost effective compared with standard treatment increased from 0.49 to 0.67 (Fig. 4) at £100000 per additional QALY gained. Hence, as would be expected, the model results are reasonably sensitive to the cost of docetaxel.

5. Discussion

This paper has illustrated how Markov models may be implemented by using Bayesian methods, in particular Gibbs sampling MCMC methods, in the software package WinBUGS (Spiegelhalter *et al.*, 1999). This environment permits great flexibility in model specification and provides the ability to perform *all* the analysis, including each individual meta-analysis used to propagate parts of the Markov model, within a unified coherent framework. The advantages that such an approach have over the equivalent classical approach include

- (a) the incorporation of greater parameter uncertainty by allowing for the fact that both the overall population effect μ and between-study precision τ^2 in the meta-analyses have both been estimated by the data,
- (b) the ability to make direct probability statements and thus direct answers to the question of interest (for example Bayesian meta-analysis can give a probability that the effect is above (or below) a particular value),
- (c) the incorporation of expert opinion either directly, as demonstrated here, or regarding the relative credibility of different data sources (Sutton and Abrams, 2001) and
- (d) that it uses the actual posterior distributions from the meta-analyses as opposed to making assumptions of normality (or some other parametric distributional form) that are necessary for the classical analysis.

Differences in parameter estimates obtained by using classical and Bayesian methods, where the latter uses vague prior distributions, will generally be small especially when the number of studies in a meta-analysis is large. However, in the above analysis, the number of studies in each meta-analysis was small (i.e. four studies or fewer) leading to quite large differences in the results from the two methods due to the considerable uncertainty in the estimation of the between-study variance parameters accounted for in the Bayesian model. This led to the greater dispersion of estimates that is seen in Fig. 3.

One limitation of the analysis that is presented above is that the same studies contribute to many of the parameter estimates which inform the decision model. It is assumed that the evidence on each parameter is 'independent'. However, an improvement would be to treat the data as truly multivariate, possibly utilizing more sophisticated meta-analysis models for multiple outcomes (Berkey *et al.*, 1998; DuMouchel, 1998). There is a similar limitation with the informative prior information where the same oncologists contribute to multiple model parameters (see O'Hagan (1998)). It should be noted that correlations are induced in the model by its

structure; for example, the estimated costs and effects are correlated within each cycle of the Gibbs sampler.

The feasibility of including informative prior distributions in the Bayesian analysis has been demonstrated. In this example including this information had relatively little effect on the overall conclusions; however, if more informative prior distributions were placed on model parameters this could have had a considerable effect on the results. Further, the use of sensitivity analyses to explore fully the effect of the prior distributions, whether deliberately informative or not (Lambert *et al.*, 2001), on the overall conclusions has been recommended (Spiegelhalter *et al.*, 2000).

Fig. 3 and Table 3 clearly show the great degree of uncertainty in the evaluation of incremental costs and benefits, making it difficult to draw conclusions and to make policy recommendations from the analysis as it stands. This is because the evidence base regarding the use of taxanes in advanced breast cancer is small. It would be desirable to be able to make recommendations for future specific research that would enable this uncertainty to be reduced. How to do this optimally is a difficult but interesting challenge. The implementation of decision theoretic methodology such as the use of the expected value of information (Claxton *et al.*, 2000; Claxton and Posnett, 1996) could be used to identify the cost–utility of conducting future research, and in particular that which would reduce uncertainty regarding specific parameters or inputs in the Markov model, and is the subject of on-going research.

The Markov model that is developed above provided a simplified representation of the process of second-line treatment for advanced breast cancer. However, in all decision modelling a balance must be maintained between modelling the real situation with sparse unreliable data and simplifying the model to fit the data that are available, as was the case here. Keeler (1996) stated that

‘... even simple models that predict outcomes can substantially aid intuition in judging cost-effectiveness of drugs with long term effects’

and that

‘... analysts must use common sense to decide which factors cannot be excluded and, as a result, how many states to allow’.

An approach for addressing the uncertainty in the structural form of the model from a Bayesian perspective has been proposed by Draper (1995) but more formal systematic approaches to exploring the sensitivity of the results to the model specification are required and this is the subject of further research. The model that is discussed in this paper was developed to illustrate the application of Bayesian methods to decision analytical modelling and focused on the comparison of doxorubicin with docetaxel. However, with a number of current treatment options available for advanced breast cancer, it is plausible for the model to be extended to incorporate more comparators.

A threat to the validity of any synthesis of evidence is the potential of publication bias (Song *et al.*, 2000), as statistically significant or positive findings are more likely to be published than negative findings. A meta-analysis based on a biased sample of the total literature will produce biased answers and likewise economic evaluations based on such syntheses will themselves be biased (Freemantle and Mason, 1997). Methods, both Bayesian and classical, exist which assess for the presence and even adjust for publication bias in a meta-analysis (Sutton *et al.*, 2000b), the simplest of which is visual assessment via a funnel plot of the precision (the inverse of the standard error) against treatment effect. A further consideration is the synthesis of different sources of evidence to inform a decision model and the fact that these different

sources could be affected by differential amounts and types of publication bias (Sutton *et al.*, 2002).

In the example that is presented throughout this paper, the posterior distribution for each model parameter that is obtained from the meta-analyses is utilized in the probabilistic decision model. However, sometimes the use of the posterior predictive distribution may be more appropriate, e.g. when decisions or inferences are of interest at the individual (unit) level rather than the population (average) level (Higgins, 2001). An example would be if we were interested in the effect of introducing taxanes for an individual hospital (i.e. unit level) as opposed to national or international level. Using the predictive distribution can considerably increase the level of uncertainty in the analysis. It is easy to estimate the posterior predictive distribution within WinBUGS and this is another advantage of the Bayesian approach to decision modelling, as this would be difficult to obtain from a classical approach.

It has been recommended in guidelines (Moher *et al.*, 1994) that when reporting systematic reviews and meta-analysis the methods that were used should be transparent and sufficient details should be reported to enable the work to be reproducible by others. In describing the Markov model that is developed here, in the same spirit, we have attempted to make all stages of the analysis transparent and reproducible. This is important since it is necessary for a full appraisal and critique of the model and facilitates future developments. Completely transparent reporting is challenging since there are a large number of components to the model (and assumptions made), to which a wide array of evidence contributes. It is our experience that often such models are not reported in sufficient detail to allow replication or critical appraisal, with many aspects of decision models being treated as a 'black box'. In addition to the guidelines of 'good modelling practice' in the economic modelling of health technologies (Akehurst *et al.*, 2000), guidelines to assist with clear and comprehensive reporting of decision models in medicine and related areas, in a similar spirit to those already developed for other types of studies (Altman *et al.*, 2001; Moher *et al.*, 1994; Stroup *et al.*, 2000), would be of great value to the promotion of evidence-based practice.

As highlighted previously in this paper, many assumptions are required to specify and evaluate a complex decision model. Although we have made an attempt to assess some of the assumptions that were made, methodology to enable a more comprehensive assessment to be carried out, considering all parameter and structural uncertainty, would be desirable.

A copy of the WinBUGS program and data that were used in the analysis is available at

<http://www.blackwellpublishing.com/rss>

Acknowledgements

The authors thank Michael Borowitz, Ruth Brown and John Hutton of MEDTAP International for their assistance with the elicitation process, and John Hutton for advice regarding earlier economic models. The authors also thank Albert Bengiat, University Hospitals of Leicester National Health Service Trust, and Kate Smith, Clatterbridge Centre for Oncology National Health Service Trust, for clinical advice regarding the development of the model described, David R. Jones, Department of Epidemiology and Public Health, University of Leicester, for comments on earlier drafts of this paper, and the two referees and Joint Editor for their valuable comments which improved the paper considerably.

NJC is funded by University Hospitals of Leicester National Health Service Trust. KRA has acted as a paid consultant to MEDTAP International on a project commissioned by Aventis (formerly Rhône-Poulenc Rorer) who manufactures Taxotere®.

References

- Abrams, K. R., Ashby, D. and Errington, R. D. (1994) Simple Bayesian analysis of a clinical trial—a tutorial. *Contr. Clin. Trials*, **14**, 255–260.
- Akehurst, R., Anderson, P., Brazier, J., Brennan, A., Briggs, A., Buxton, M., Cairns, J., Calvert, N., Claxton, K., Dixon, S., Fryback, D., Gallivan, S., Green, C., Lloyd, A., McCabe, C., Mitchell, A., Nicholl, J., O'Brien, B., Roberts, J., Sculpher, M., Severens, H., Sullivan, S. and Veenstra, D. (2000) Decision analytical modelling in the economic evaluation of health technologies: consensus conference on guidelines on economic modelling in health technology assessment. *Pharmacoeconomics*, **17**, 443–444.
- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gotzsche, P. C. and Lang, T. (2001) The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann. Intern. Med.*, **134**, 663–694.
- Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F. and Colditz, G. A. (1998) Meta-analysis of multiple outcomes by regression with random effects. *Statist. Med.*, **17**, 2537–2550.
- Bonnetterre, J., Roche, H., Monnier, A., Serin, D., Fargeot, P., Guastalla, J. P., Rios, M., Namer, M., Culine, S., Tubiana, M., Assadourian, S. and Gaudin, I. (1997) Taxotere (TXT) versus 5-fluorouracil + navelbine (FUN) as second-line chemotherapy (CT) in patients (pts) with metastatic breast cancer (MBC). *Proc. ASCO*, **16**, A564.
- Bontenbal, M., Andersson, M., Wildiers, J., Cocconi, G., Jassem, J., Paridaens, R., Rotmensz, N., Sylvester, R., Mouridsen, H. T., Klijn, J. G. M. and van Oosterom, A. T. (1998) Doxorubicin vs epirubicin, report of second-line randomized phase III/III study in advanced breast cancer. *Br. J. Cancer*, **77**, 2257–2263.
- Brennan, A. and Akehurst, R. (2000) Modelling in health economic evaluation: what is its place?; what is its value? *Pharmacoeconomics*, **17**, 445–459.
- Briggs, A. H. (2000) Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics*, **17**, 479–500.
- Briggs, A. and Sculpher, M. (1998) An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics*, **13**, 397–409.
- Brown, R. E. and Hutton, J. (1998) Cost-utility model comparing docetaxel and paclitaxel in advanced breast cancer patients. *Anti-cancer Drugs*, **9**, 899–907.
- Cancer Research Campaign (1996) *Breast Cancer—Fact Sheet*. London: Cancer Research Campaign.
- Carlin, B. P. and Louis, T. A. (2000) *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edn. London: Chapman and Hall.
- Chan, S., Friedrichs, K., Noel, D., Pinter, T., Van Belle, S., Vorobiof, D., Duarte, R., Gil, M. G., Bodrogi, I., Murray, E., Yelle, L., von Minckwitz, G., Korec, S., Simmonds, P., Buzzi, F., Mancha, R. G., Richardson, G., Walpole, E., Ronzoni, M., Murawsky, M., Alakl, M., Riva, A. and Crown, J. (1999) Prospective randomized trial of Docetaxel versus Doxorubicin in patients with metastatic breast cancer. *J. Clin. Oncol.*, **17**, 2341–2354.
- Claxton, K., Lacey, L. F. and Walker, S. G. (2000) Selecting treatments: a decision theoretic approach. *J. R. Statist. Soc. A*, **163**, 211–225.
- Claxton, K. and Posnett, J. (1996) An economic approach to clinical trial design and research priority-setting. *Hlth Econ.*, **5**, 513–524.
- Cooper, N. J., Sutton, A. J. and Abrams, K. R. (2002) Decision analytical economic modelling within a Bayesian framework: application to prophylactic antibiotics use for caesarean section. *Statist. Meth. Med. Res.*, to be published.
- Department of Health (1999) NHS reference costs. (Available from www.doh.gov.uk/nhsexec/refcosts.htm.)
- DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Contr. Clin. Trials*, **7**, 177–188.
- Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *J. R. Statist. Soc. B*, **57**, 45–97.
- DuMouchel, W. (1998) Repeated measures meta-analysis. *Bull. Int. Statist. Inst.*, **57**, book 1, 285–288.
- Feeny, D., Furlong, W., Boyle, M. and Torrance, G. W. (1995) Multi-attribute health status classification systems: health utilities index. *Pharmacoeconomics*, **7**, 490–502.
- Fornier, M., Munster, P. and Seidman, A. D. (1999) Update on the management of advanced breast cancer. *Oncology*, **13**, 647–658.
- Freemantle, N. and Mason, J. (1997) Publication bias in clinical trials and economic analyses. *Pharmacoeconomics*, **12**, 10–16.
- Furlong, W., Feeny, D. and Torrance, G. W. (1990) *Guide to Design and Development of Health State Utility Instrumentation*. Hamilton: McMaster University.
- Genest, C. and Zidek, J. V. (1986) Combining probability distributions: a critique and an annotated bibliography. *Statist. Sci.*, **1**, 114–148.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds) (1996) *Interdisciplinary statistics. In Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Gore, S. M. (1987) Biostatistics and the Medical Research Council. *Med. Res. Coun. News*, 19–20.
- Hardy, R. J. and Thompson, S. G. (1996) A likelihood approach to meta-analysis with random effects. *Statist. Med.*, **15**, 619–629.
- Higgins, J. (2001) How do we interpret random effects meta-analyses. *Workshp Meta-analysis, London*.

- Hutton, J., Brown, R., Borowitz, M., Abrams, K., Rothman, M. and Shakespeare, A. (1996) A new decision model for cost-utility comparisons of chemotherapy in recurrent metastatic breast cancer. *Pharmacoeconomics*, **9**, 8–22.
- Keeler, E. (1996) Decision trees and Markov models in cost-effectiveness research. In *Valuing Health Care* (ed. F. A. Sloan), pp. 185–205. Cambridge: Cambridge University Press.
- Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. and Jones, D. R. (2001) How vague is vague? *22nd A. Conf. International Society for Clinical Biostatistics, Stockholm*, abstr.
- Lilford, R. and Royston, G. (1998) Decision Analysis in the selection, design and application of clinical and health services research. *J. Hlth Serv. Res. Poly.*, **3**, 159–166.
- Lister-Sharp, D., McDonagh, M., Khan, K. S. and Kleijnen, J. (2000) A systematic review of the effectiveness and cost-effectiveness of the taxanes used in the treatment of advanced breast and ovarian cancer. *National Health Service Research and Development Health Technology Assessment Programme*.
- Miller, D. K. and Homan, S. M. (1994) Determining transition probabilities: confusion and suggestions. *Med. Decsn Mak.*, **14**, 52–58.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D. and Stroup, D. F. (1994) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet*, **354**, 1896–1900.
- Nabholtz, J. M., Senn, H. J., Bezwoda, W. R., Melnychuk, D., Deschenes, L., Douma, J., Vandenberg, T. A., Rapoport, B., Rosso, R., Trillet-Lenoir, V., Drbal, J., Molino, A., Nortier, J. W. R., Richel, D. J., Nagykalnai, T., Siedlecki, P., Wilking, N., Genot, J. Y., Hupperets, P. S. G. J., Pannuti, F., Skarlos, D., Tomiak, E. M., Murawsky, M., Alakl, M., Riva, A. and Aapro, M. (1999) Prospective randomized trial of Docetaxel versus Mitomycin plus Vinblastine in patients with metastatic breast cancer progressing despite previous anthracycline-containing chemotherapy. *J. Clin. Oncol.*, **17**, 1413–1424.
- Natarajan, R. and Kass, R. E. (2000) Reference Bayesian methods for generalised linear mixed models. *J. Am. Statist. Ass.*, **95**, 227–237.
- O'Brien, M. E. R., Leonard, R. C., Barrett-Lee, P. J., Eggleton, P. H. and Bizzari, J. P. (1999) Docetaxel in the community setting: an analysis of 377 breast cancer patients treatment with docetaxel (Taxotere) in the UK. *Ann. Oncol.*, **10**, 205–210.
- O'Hagan, A. (1998) Eliciting expert beliefs in substantial practical applications. *Statistician*, **47**, 21–35.
- Parmar, M. K., Griffiths, G. O., Spiegelhalter, D. J., Souhami, R. L., Altman, D. G. and van der Scheuren, E. (2001) Monitoring of large randomised clinical trials: a new approach with Bayesian methods. *Lancet*, **358**, 375–381.
- Parmar, M. K., Spiegelhalter, D. J. and Freedman, L. S. (1994) The CHART trials: Bayesian design and monitoring in practice. *Statist. Med.*, **13**, 1297–1312.
- Parmigiani, G., Samsa, G., Ancukiewicz, M., Lipscomb, J., Hasselblad, V. and Matchar, D. B. (1997) Assessing uncertainty in cost-effectiveness analyses. *Med. Decsn Mak.*, **17**, 390–401.
- Scottish Health Purchasing Information Centre (1997) *Breast Cancer*. Aberdeen: Scottish Health Purchasing Information Centre.
- Sjöström, J., Blomqvist, C., Mouridsen, H., Pluzanska, A., Ottosson-Lönn, S., Bengtsson, N. O., Østenstad, B., Mjaaland, I., Palm-Sjövall, M., Wist, E., Valvere, V., Anderson, H. and Bergh, J. (1999) Docetaxel compared with sequential methotrexate and 5-fluorouracil in patients with advanced breast cancer after anthracycline failure: a randomised phase III study with crossover on progression by Scandinavian breast cancer group. *Eur. J. Cancer*, **35**, 1194–1201.
- Song, F., Easterwood, A., Gilbody, S., Duley, L. and Sutton, A. J. (2000) Publication and other selection biases in systematic reviews. *Hlth Technol. Assessmnt*, **4**.
- Sonnenberg, F. A. and Beck, J. R. (1993) Markov models in medical decision making: a practical guide. *Med. Decsn Mak.*, **13**, 322–338.
- Spiegelhalter, D. J., Myles, J. P., Jones, D. R. and Abrams, K. R. (2000) Bayesian methods in health technology assessment: a review. *Hlth Technol. Assessmnt*, **4**.
- Spiegelhalter, D. J., Thomas, A. and Best, N. G. (1999) *Win BUGS Version 1.2 User Manual*. Cambridge: Medical Research Council Biostatistics Unit.
- Stinnett, A. A. and Mullahy, J. (1998) Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med. Decsn Mak.*, **18**, suppl., S68–S80.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A. and Thacker, S. B. (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. *J. Am. Med. Ass.*, **283**, 2008–2012.
- Sutton, A. J. and Abrams, K. R. (2001) Bayesian methods in meta-analysis and evidence synthesis. *Statist. Meth. Med. Res.*, **10**, 277–303.
- Sutton, A. J., Abrams, K. R. and Jones, D. R. (2002) Generalised synthesis of evidence and the threat of dissemination bias: the example of electronic fetal heart rate monitoring. *Technical Report 02-01*. Department of Epidemiology and Public Health, University of Leicester, Leicester.
- Sutton, A. J., Abrams, K. R., Jones, D. R. and Sheldon, T. A. S. F. (2000a) *Methods for Meta-analysis in Medical Research*. Chichester: Wiley.
- Sutton, A. J., Song, F., Gilbody, S. M. and Abrams, K. R. (2000b) Modelling publication bias in meta-analysis: a review. *Statist. Meth. Med. Res.*, **9**, 421–445.