
A. Briggs, R. Goeree, G. Blackhouse, B. O’Brien
CHEPA WORKING PAPER SERIES

The Centre for Health Economics and Policy Analysis (CHEPA) Working Paper Series provides for the circulation on a pre-publication basis of research conducted by CHEPA faculty, staff, and internal associates. The Working Paper Series is intended to stimulate discussion on analytical, methodological, quantitative, and policy issues in health economics and health policy analysis. The views expressed in the papers are the views of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of Working Papers are encouraged to contact the author(s) with comments, criticism, and suggestions.

A list of papers in the Series and papers in the CHEPA Health Policy Commentary Series, together with information on how to obtain specific titles, is provided at the end of this paper. Further information can be obtained by contacting the Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, L8N 3Z5; (905) 525-9140, Ext. 22135.

THE CENTRE FOR HEALTH ECONOMICS AND POLICY ANALYSIS

The Centre for Health Economics and Policy Analysis officially opened in May, 1988. The objectives of the Centre include research, education, and the promotion of interactions among academic researchers and policy makers. CHEPA includes researchers from a number of departments in the University including Clinical Epidemiology and Biostatistics, Management Science, Family Medicine, and Geography.
Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for Gastro-Esophageal Reflux Disease

Andrew H. Briggs

Health Economics Research Centre
University of Oxford

Ron Goeree
Gord Blackhouse
Bernie O'Brien

Centre for Evaluation of Medicines
McMaster University

ABSTRACT

When choosing between mutually exclusive treatment options, it is common to construct a cost-effectiveness frontier on the cost-effectiveness plane that represents efficient points from among the treatment choices. Treatment options internal to the frontier are considered inefficient and are excluded either by strict dominance or by appealing to the principle of extended dominance. However, when uncertainty is considered, options excluded under the baseline analysis may form part of the cost-effectiveness frontier. By adopting a Bayesian approach, where distributions for model parameters are specified, uncertainty in the decision concerning which treatment option should be implemented is addressed directly. The approach is illustrated using an example from a recently published cost-effectiveness analysis of different possible treatment strategies for gastro-esophageal reflux disease. It is argued that probabilistic analyses should be encouraged since they have potential to summarise the strength of evidence in favour of particular treatment choices.
ACKNOWLEDGEMENTS

Participants at the workshop on “Health Economics and Outcomes Research: Making Statistics More Relevant to Decision-Makers” held in Washington, DC, September 1998 and at a Harvard Centre for Risk Analysis seminar, in particular Denny Fryback, Karl Claxton and Milton Weinstein, for discussions concerning the issues addressed in this paper. Of course, all errors are our own responsibility.
1. Introduction

It is now increasingly common for economic evaluations to be conducted alongside clinical trials. Recent research attention has been focussed on how to handle uncertainty in these so-called stochastic cost-effectiveness analyses where patient level data are available on the costs and effects of treatment options (O’Brien et al. 1994; van Hout et al. 1994; Stinnett and Paltiel, 1997). However, the majority of economic evaluations still employ a decision analytic modelling framework to synthesise data from a number of sources (Briggs and Gray, 1999). Such cost-effectiveness models are often described as deterministic analyses. Although the limitations of simple univariate sensitivity analysis are well known, this remains the most popular technique for handling uncertainty in cost-effectiveness models.

Probabilistic sensitivity analysis is an alternative approach which involves specifying distributions for input parameters in the model and employing Monte Carlo simulation to sample from these distributions allowing the joint effect of parameter uncertainty to be assessed (Critchfield et al. 1986; Doubilet et al. 1985). A number of commentators have suggested that probabilistic sensitivity analysis methods should be used to handle uncertainty in cost-effectiveness models (O’Brien et al. 1994; Briggs, 2000), including the U.S. panel on cost-effectiveness analysis (Manning et al. 1996). Despite these recommendations, few probabilistic analyses of cost-effectiveness models have been undertaken. The relative paucity of probabilistic analyses may be due to the increased complexity of the approach and a lack of clarity concerning which distributions for input parameters are appropriate. The aim of this paper is to demonstrate the use of probabilistic sensitivity analysis to handle uncertainty in a cost-effectiveness decision problem relating to alternative treatment options for gastro-esophageal reflux disease (GERD). We argue that adopting a Bayesian approach to uncertainty offers both technical and conceptual advantages over traditional sensitivity analyses. It is natural to interpret uncertainty in the input parameters in a Bayesian way reflecting our belief that the parameter could take different values, rather than using the standard frequentist notion that parameters have a single true value that does not vary. We believe that employing a Bayesian approach will encourage analysts to consider very carefully uncertainty in parameter values, which should lead to better quality analyses. A Bayesian approach also allows a more intuitive interpretation of probability – in particular, we show how the study question of whether a treatment is cost-effective can be answered directly in the form of a probability that the intervention is cost-effective. Furthermore, we demonstrate this approach in the case of multiple treatment options for GERD, rather than the standard two alternative treatment approach that is the norm in the majority of analyses.
The paper is structured as follows. In the next section we give a brief introduction to the decision problem, the structure of the model and the results of the previously published deterministic cost-effectiveness model. The section that follows then demonstrates how the model can be made probabilistic by specifying distributions for the input parameters following standard principles of Bayesian methods. The assumptions and calculations involved in specifying these distributions are laid out in full. Results of the probabilistic analysis are then presented on the cost-effectiveness plane and summarised through the use of cost-effectiveness acceptability curves. These results, and the general probabilistic approach to cost-effectiveness modelling are discussed in the final section of the paper.
2. Treatment strategies for gastro-esophageal reflux disease

In this section, a model for assessing the cost-effectiveness of six management strategies for the treatment of gastro-esophageal reflux disease is briefly outlined. Full details of the model have been presented in detail in a previous publication (Goeree et al. 1999). First the structure and assumptions concerning the decision model are discussed. Second the results of the deterministic analysis are presented. Finally, the limitations of the originally reported univariate sensitivity analysis are highlighted.

2.1 A model for assessing the cost-effectiveness of GERD treatment

Gastro-esophageal reflux disease (GERD) is a common condition that results from regurgitation of acid from the stomach into the esophagus. The most frequent symptom of GERD is heartburn and the majority of patients with GERD require pharmacotherapy to reduce acid secretion. Currently, the choice of first-line antisecretory therapy is between the $H_2$-receptor antagonists ($H_2$RAs), such as ranitidine and cimetidine and proton pump inhibitors (PPIs) such as omeprazole. Although they have higher acquisition costs, PPIs have been found to be more efficacious than $H_2$RAs in terms of both the rate and speed of healing (Chiba et al. 1993; Chiba et al. 1993).

The objective of the original study was to compare, over a one-year period, the expected costs and outcomes of alternative drug treatment strategies for the management of patients with erosive esophagitis confirmed by endoscopy, but without complications such as Barrett’s esophagus or stricture. Outcomes are quantified in terms of GERD recurrence and weeks per year without GERD as indicated by data from clinical trials on healing and recurrence of esophagitis.

Treatment strategies and model structure

Six strategies involving different combinations of first-line agents and change of therapy conditional on failure to heal or recurrence of GERD were modelled.

Strategy A: Intermittent PPI. Acute treatment with a PPI for 8 weeks and then no further treatment with prescription medication until recurrence.

Strategy B: Maintenance PPI. Acute treatment with a PPI for 8 weeks then continuous maintenance treatment with a PPI (same dose).

Strategy C: Maintenance $H_2$RA. Acute treatment with an $H_2$RA for 8 weeks and then continuous maintenance treatment with an $H_2$RA (same dose).
Strategy D: Step-Down Maintenance Prokinetic Agent. Acute treatment with a prokinetic agent (PA) for 12 weeks and then continuous maintenance treatment with a lower dose of PA.

Strategy E: Step-Down Maintenance H$_2$RA. Acute treatment with a PPI for 8 weeks and then continuous maintenance treatment with a H$_2$RA.

Strategy F: Step-Down Maintenance PPI. Acute treatment with a PPI for 8 weeks and then continuous maintenance treatment with a lower dose PPI.

Treatment options A to F represent clinical strategies rather than single drug treatments for the management of erosive-esophagitis where the physician is assumed to increase the dose of a drug or switch to another drug if the patient fails to respond to the first-line treatment. The logic of these assumptions regarding stepping-up dosage or switching can be found in Table 1. The structure of the decision tree that was developed is shown in Figure 1 and was based on the treatment strategies and step-up switching algorithms in Table 1. The model is recursive in two 6-month periods; hence, probabilities of recurrence in the period to 12 months are conditional upon recurrence or non-recurrence in the period from 0 to 6 months.

Treatment outcomes

For GERD, the most commonly used formulation of outcome for economic evaluation has been either esophagitis-free or symptom-free time in a period of follow-up. The advantage of such a measure is that it combines two important aspects of efficacy: (i) the speed with which esophagitis is healed; and (ii) the likelihood of esophagitis recurring. In this analysis, the primary outcome measure is GERD-free time during the 12-month period of the model, defined as the time where the esophagitis is healed. A meta-analysis of healing and recurrence studies published to November 1997 was undertaken to estimate healing and recurrence probabilities together with associated GERD-free time. Full details of this analysis are given in the original study (Goeree et al. 1999).

Resource use and unit costs

Generic prices were used for drugs where a generic equivalent is available, employing the ‘best available price’ from the Ontario Drug Benefit (ODB) program (Ministry of Health, 1996) together with a 10% pharmacy mark-up charge. A dispensing fee of Can $4.11 was used (i.e. ODB Program fee of Can $6.11 less a Can $2.00 patient co-payment).

Cost estimates for physician fees were taken from the Physician fee schedule for Ontario (Ontario Ministry of Health, 1992) and procedure costs, such as endoscopy, were estimated
from a hospital participating in the Ontario Case Costing Project in Southwestern Ontario (Ontario Hospital Association, 1995).

To estimate the costs associated with the management of patients with symptoms of GERD recurrence, information on clinical practice patterns and resource utilisation was obtained by convening an expert physician panel and using a modified Delphi technique (Park et al. 1986). Estimated resource utilisation was then combined with unit cost information to give the average cost associated with each recurrence under each management strategy.

2.2 Results of the deterministic CEA

The decision tree model outlined above was evaluated to estimate the expected costs and the expected weeks without GERD in the 12-month period of the model. The analysts of the original study took the accepted conventional approach to examining the cost-effectiveness of the alternative strategies (Weinstein, 1995; Gold et al. 1996; Drummond et al. 1997). First, it was determined if any strategies were dominated by other strategies having both lower costs and greater therapeutic effects. Second, it was determined if any strategies were dominated through the principles of extended dominance (i.e. whether linear combinations of other strategies can produce the same (or greater) benefit at lower (or the same) cost (Cantor, 1994). Finally, among non-dominated treatment options, incremental cost-effectiveness ratios were calculated by comparing each option to the next most costly and effective intervention. This process produces an ‘efficient frontier’ of increasingly more costly, but more effective strategies. The results of the analysis is presented on the cost-effectiveness (CE) plane in Figure 2, which also shows the efficient frontier.

The figure clearly shows that step down maintenance PA (strategy D) is dominated by maintenance H₂RA, (strategy C), intermittent PPI (strategy A) and step-down maintenance H₂RA (strategy E). The efficient frontier is given by the lines joining strategies C (the origin) A, E and B. Strategy F is internal to this frontier indicating that it is extended dominated by a linear combination of strategies E and B. The slope of the frontier reflects incremental cost-effectiveness – the additional cost at which additional effects can be purchased.

2.3 Limitations of conventional sensitivity analysis

The three-step approach to generating the efficient frontier described in the section above was undertaken deterministically without consideration of uncertainty. Of course, parameter estimates employed in the model are not known with certainty and it is important to explore the implications of parameter uncertainty for the results of the analysis. In particular, given the
importance of excluding strictly dominated and extended dominated interventions from the analysis before calculating the frontier, the extent to which particular strategies are part of (or can be excluded from) the frontier should be assessed.

In the original analysis, the authors examined the effect of a number of the parameters in their model using conventional sensitivity analysis techniques. They showed how the uncertainty in the parameter values chosen for the baseline analysis might affect the frontier. The sensitivity analysis that the authors presented was much less arbitrary than that of many reported cost-effectiveness analyses since the outcome ranges chosen were the 95% confidence limits from the reported meta-analyses of healing and recurrence rates. The authors reported that

“there were marked differences in expected costs, recurrences and weeks with GERD when using the lower and upper CIs for both healing and recurrence rates. However, there were no changes in the relative ranking of strategies for either costs or outcomes. The basic conclusions of the base-case analysis were not altered by using the lower or upper 95% CIs for healing or recurrence rates.” (Goeree et al. 1999, p689 & 671.)

Despite this convincing argument, we might still be concerned that the full effects of uncertainty are more important that the authors suggest. It is well known that conventional univariate sensitivity analysis, whereby individual parameters are varied while maintaining all remaining parameters at their baseline value, is likely to underestimate uncertainty since, in reality, parameters will not vary in isolation.
3. A Bayesian approach to probabilistic sensitivity analysis

In this section, the general appeal of adopting a Bayesian approach to probabilistic analysis of cost-effectiveness models is presented. Particular emphasis is given to the choice of distributions for the different types of parameters commonly encountered in cost-effectiveness models. In section 4, the Bayesian approach to probabilistic analysis of cost-effectiveness models is illustrated from the GERD example.

3.1 Probabilistic analysis

Probabilistic sensitivity analysis involves specifying distributions for model parameters to represent uncertainty in their estimation and employing Monte Carlo simulation to select values at random from those distributions. In this way, probabilistic models allow the effects of joint uncertainty across all the parameters of the model to be considered. Note that for standard frequentist analyses (such as practised in almost all clinical trials) parameters to be estimated from the data are considered to have true values that do not vary. Probabilities attached to confidence limits relate to the long-run coverage probabilities of the intervals if the same experiment were to be repeated many times.

By contrast, in probabilistic modelling, parameters are considered random variables, which can take a range of values described by the specified distribution. Although these distributions will represent ‘degrees of belief’ in the parameters of interest, it does not necessarily follow that the analysis will become automatically ‘subjective’ (the great fear of many of those who object to Bayesian methodology). Where data are lacking and it becomes necessary to engage experts to provide information on prior distributions then a number of experts should be consulted in order that the distributions reflect uncertainty between experts rather than representing the subjective beliefs of a single expert. Eddy and colleagues have outlined just such an approach to synthesising data based on Bayesian methods that they term the ‘confidence’ profile technique (Eddy et al. 1990; Eddy et al. 1990).

3.2 Choosing distributions for the parameters

Parameters in decision models represent summary values related to the average experience across a population of (potential) patients. Therefore the relevant uncertainty to capture in the formation of a distribution for the parameter is second-order uncertainty related to the sampling distribution of the parameter, not the variability in the values observed in a particular population (first-order uncertainty) (see Stinnett and Paltiel, 1997 for further discussion). Although an assumption of normality for parameters is widely used in statistics it is worth remem-
bering that the assumption is based on asymptotics (the *central limit theorem*) and that the normal distribution has no bounds on values it can take. In practice, parameters of the model will have logical limitations on the values they can take. In this section we discuss five different parameter types commonly employed in cost-effectiveness models: probabilities, resource items, unit costs, relative risks and utility scores. For each we discuss the nature of the data informing parameter estimates, the logical bounds on the parameter and the way in which Bayesian methods can help to select distributions for parameters.

**Probability parameters**

Probabilities for cost-effectiveness models are often based on the observed proportions of the event of interest, say the number of successfully treated cases. Consider that at an individual level a treated patient is either classed as a success or as a failure, therefore, the data can be considered as independent Bernoulli trials leading to a Binomial form of the data likelihood. With such data it is natural to use the proportion of successful patients as the estimate of the corresponding probability in the model. However, in considering the distribution of that probability, note that the binomial distribution is a discrete distribution related to the sample size of the study generating the data, whereas it makes sense to model the distribution of probability in the model as continuous.

Standard frequentist methods for estimating a confidence interval for a proportion involves calculating the Binomial estimate of variance and assuming a normal sampling distribution in order to generate the interval

\[ p - 1.96 \times \sqrt{\frac{p(1-p)}{n}}, \quad p + 1.96 \times \sqrt{\frac{p(1-p)}{n}} \]

(where \( p \) is the proportion and \( n \) is the sample size) (Altman, 1991). While this method gives a good approximation to the true confidence interval when \( p \) is not close to zero or one, the assumption of normality is not appropriate for probabilistic sensitivity analysis. This is because the probability is known to be bounded on the interval zero-one, while the normal distribution will (eventually) generate values outside this interval in a Monte Carlo simulation.

Fortunately, Bayesian methods provide a method for moving from the discrete Binomial likelihood to the continuous uncertainty concerning the probability parameter. The Beta distribution is a continuous distribution on the interval zero-one and is *conjugate* to the binomial distribution. This means that if it is possible to represent prior belief using a Beta distribution,\(^1\)

---

\(^1\) Fortunately, by varying the two parameters of the Beta distribution, a wide variety of possible shapes to the distribution over the interval can be obtained: skewed, symmetric, uniform, near normal and even U-shaped.
then the integration of that prior belief with the Binomial data has a closed form with the result that the posterior distribution of the probability will also follow a Beta distribution. One parameterisation of the Beta distribution Beta\( (r,n) \) is has a similar interpretation to \( r \) successes from \( n \) trials. The mean and variance of this distribution is given by

\[
\text{mean} = \frac{r}{n}, \quad \text{sd} = \sqrt{\frac{r(n - r)}{n^2(n + 1)}}.
\]

Furthermore, if we can specify a prior distribution as Beta\( (r',n'-r') \) then following an observation of \( r \) successes and \( n-r \) failures in \( n \) trials, the application of Bayes theorem yields the result that the posterior distribution is Beta\( (r'+r,n'-r'+n-r) \). Where no prior information exist as to the probability, it is appropriate to use an ‘uninformative’ or reference prior. Although Beta\( (1,1) \), which yields a uniform distribution over the interval zero-one, seems an intuitively obvious choice of prior, in fact, what consitutes ‘uninformative’ in this context is not as straightforward as it appears (Pratt et al. 1995)and Beta\( (0.5,0.5) \) which yields a U-shaped interval is also a popular choice for a reference prior. Since uninformative priors will be dominated by the data, the issue of which uninformative prior to employ is unlikely to be of practical importance where data are available to update that prior.

**Resource item parameters**

All economic analyses are concerned with the use of resources. The numbers of resource items that a patient utilises can be considered a count variable. The Poisson distribution with parameter \( \lambda \) (which gives both the mean and variance of the distribution) is often used to model count data. If we are interested in the distribution of the mean resource use for a group of patients we could use the Poisson estimate of variance to obtain a standard error for the mean resource use, relying on the central limit theorem to give a normal sampling distribution. However, this may be problematic for smaller samples due to a non-negligible probability that the normal distribution could take a value less than zero, when it is clear that mean resource use cannot be negative.

Again, the Bayesian approach provides a solution. The Gamma distribution is conjugate to the Poisson distribution, is constrained to be positive and is fully continuous. Therefore, the Gamma distribution for the mean resource use can be specified without fear of generating inconsistent values in a probabilistic analysis.
Unit cost parameters
Unit costs are applied to resource volumes in order to evaluate all resource use on a common (monetary) scale. Note that the unit of analysis for such costs is different to that for other parameters – unit costs are typically calculated across a broad group of patients. The unit cost of a surgical procedure or stay in a particular ward will typically be given at the level of the hospital (or similar provider unit). By contrast, the unit cost of a drug or device may be set nationally or provincially and may not vary at all, within the context of a country-specific cost-effectiveness analysis. Furthermore, the unit cost of a resource item is strictly continuous, unlike the data on resource use considered above. While unit costs are constrained to be positive, it is less likely that unit costs will be close to zero (in the sense of their variability rather than absolute value) than the resource items they seek to value. For this reason, it may be more appropriate to assume a normal distribution for representing uncertainty in unit cost information, although it is worth considering the nature of the unit of sampling being considered. It is perhaps telling that most economic analyses conducted alongside clinical trials treat unit costs as fixed rather than stochastic.

Relative risk parameters
It is very common for economic models to include relative risks as parameters. This mirrors the fact that relative risk is often the primary outcome in clinical trials. Methods for calculating confidence intervals for relative risk estimated in such trials assumes the central limit theorem will lead to log relative risk (which is additive) being normally distributed such that confidence intervals can be determined in the usual way. A confidence interval for relative risk is then obtained by anti-logging (exponentiating) the confidence limits on the log scale. This standard approach to confidence interval estimation clearly suggests an equivalent approach to specifying a log-normally distributed parameter for relative risk to be used in a probabilistic sensitivity analysis. Furthermore, since the normal distribution is self-conjugate (a normal prior and a normal data likelihood generate a normal posterior distribution), the application of Bayes theorem on a normally distributed parameter is especially straightforward.
4. **A probabilistic analysis of choosing treatment strategies for GERD**

In this section, we describe how a probabilistic sensitivity analysis of this decision problem was undertaken in order to more fully account for uncertainty in the choice of treatment strategy for GERD. Within the model there are three main categories of parameters. Model probabilities relating to the healing and recurrence rates of GERD symptoms, parameters relating to the level of resource consumption by patients with GERD symptoms and unit costs of those resources. Each of these parameter categories are discussed in detail below. Note that the outcome variables in the model – weeks free of GERD – are completely determined by the healing and recurrence rates and are therefore endogenous variables in the model.

4.2 **Parameter distributions**

*Distributions for the healing and recurrence probabilities*

All patients begin the model with GERD. Following first-line therapy there is a probability that their GERD will have healed. Once GERD has healed there is then the probability that it will recur. The healing and recurrence probabilities were estimated from the literature. Consider that at an individual level a patient with GERD has either healed/experienced a recurrence or not. At the individual level, therefore, the data from clinical investigation of healing and recurrence can be considered as independent Bernoulli trials leading to a Binomial form of the data likelihood as described above -- hence a beta distribution was chosen to represent uncertainty in this parameter.

The original study went to some lengths to present a rigorous meta-analysis of healing and recurrence probabilities. This method resulted in estimates of constant hazards for healing probabilities and estimates of proportions of patients recurring in two periods: 0-6 months following healing and 6-12 months following healing, together with associated estimates of standard error. Equating the mean and standard errors from the meta-analysis to the estimates of mean and standard error of the beta distribution given above allowed the Beta distributions to be fitted directly. Due to the random-effects assumption, this is more conservative than the direct Bayesian updating approach described above. Details of the distributions for the healing hazards and the recurrence probabilities fitted by this method are given in Table 3.
**Distributions for resource use assumptions**

In contrast to the rigorous meta-analytic approach employed to summarise the wealth of information on the healing and recurrence rates associated with different drug interventions for GERD, the information on resource use, particularly the level of investigations received by patients following a recurrence, were extremely sparse. Although in the original study a Delphi panel of experts was convened in order to estimate the likely experience of patients, the purpose of the panel was to forge consensus and no information on the variance of estimates that emerged prior to consensus of the experts remains. Therefore the assumptions concerning the distributions of estimated resource use are much more arbitrary.

For the estimated number of visits to general practitioners and for endoscopic investigation a gamma distribution was assumed. This is because the number of visits is constrained to be positive and the gamma distribution is only defined for positive values. Again a method of moments approach to fitting was employed such that the mean of the gamma distribution was equal to the point estimates of the visits generated by the expert panel and assuming the standard error was half that value (i.e. assuming the coefficient of variation was 0.5).

For the proportions of patients receiving the various investigative procedures, it was assumed that the expert panel had related their estimates to a hypothetical cohort of 100 patients. Therefore, a beta distribution was again employed as if the event rates given by the expert panel were per one hundred. Given the considerable experience of the panel with GERD treatment, it is likely that this approach is conservative.

It is assumed that variation in medication use is negligible, such that all patients obtain their prescriptions and all prescriptions accord with the treatment strategies under evaluation. The chosen distributions for the resource use parameters are presented in Table 4.

**Variation in unit costs of resources**

The second component of variation in cost is the potential variation in unit cost estimates. We do not believe that it is appropriate to handle variation in drug prices probabilistically since, at the point of the evaluation, drug prices are determined by the manufacturers. Of course, there may be some uncertainty concerning which drugs it is appropriate to prescribe, but that is part of the decision problem and is best handled outside of the probabilistic component of the analysis. Therefore drug prices were not varied in this analysis.
A separate issue relates to the use of scheduled information of the cost of resource items in Ontario, Canada, where the original study was carried out. Although there is certainly an issue concerning whether scheduled reimbursement values for resources reflect the true opportunity cost of those resources, it is not clear how such uncertainty could be represented in this model. Therefore, all unit costs were taken as being deterministic and were not ascribed distributions in this analysis. While there are some problems with this approach (Rittenhouse et al. 1999), this is not conceptually different to the approach taken in stochastic CEA alongside clinical trials where it is typical for unit costs to be treated as fixed.

4.3 Results of the probabilistic analysis

Having specified distributions for all the relevant parameters of the model, the probabilistic analysis was undertaken by randomly sampling from each of the parameter distributions and calculating the expected costs and expected weeks free of GERD for that combination of parameter values. This process formed a single replication of the model results and a total of 10,000 replications were performed in order to examine the distribution of the resulting cost and outcomes for each strategy. The results of these 10,000 replications from the model are presented on the CE plane in Figure 3 together with the baseline estimate of the efficient frontier.

It is clear that for each of the individual replications an efficient frontier could be calculated together with the incremental cost-effectiveness ratios for treatments on the frontier. In particular, Figure 3 shows how it may not be possible to rule out strategy F, the strategy based on step-down maintenance PPI, since it potentially forms part of the frontier in many replications. Having generated the estimated joint distribution of costs and effects for each of the treatment strategies, how should this information be summarised? While the recent health economic literature has widely explored the handling of uncertainty in the case of two mutually exclusive treatment alternatives (refs), there has been little work on handling uncertainty in the presence of many mutually exclusive treatment options.

If the shadow price (or the maximum that society is willing to pay) for a week free of GERD symptoms was known, it would be possible to choose between all of the treatment strategies, not just identify those that form the efficient frontier. Therefore, conditional on knowing that shadow price there is only one treatment of choice from the six strategies under evaluation and the proportion of times that an intervention is the treatment of choice from the 10,000 replications of the model gives the strength of evidence in favour of that treatment. Although it would be possible to identify the efficient frontier, calculate the incremental cost-effectiveness ratios and choose one strategy from the six available for each of the 10,000 replications, a much more straightforward approach exists.
The net-benefit framework has been argued to offer many advantages for handling uncertainty in cost-effectiveness analysis (Stinnett and Mullahy, 1998) and overcomes the particular problem associated with negative ICERs (Stinnett and Mullahy, 1997). A further property is that while average cost-effectiveness ratios have no meaningful interpretation, average net-benefits have the useful property that the incremental net-benefit between any two treatments can be calculated from the difference between their individual average net-benefits (Stinnett and Mullahy, 1998). Therefore, the treatment of choice from the six strategies under evaluation will be the treatment with the greatest average net-benefit. This must be the case since only that treatment will have a positive incremental net-benefit when compared to any other treatment alternative. Therefore, the proportion of times a strategy has the highest net-benefit among the 10,000 replications of the model gives the strength of evidence in favour of that strategy being cost-effective.

Of course, in reality the shadow price of a week free of GERD symptoms is not known. However, by plotting out, for all possible values of the shadow price, the proportion of times the intervention has the greatest net-benefit much can be learned concerning the implications of the estimated uncertainty for the treatment decision. Figure 4 shows the result of just such an exercise for the probabilistic evaluation of the GERD model presented in Figure 3. These curves are conceptually the same as the use of acceptability curves to summarise uncertainty on the CE plane in a two-treatment decision problem (van Hout et al. 1994).
5. Discussion

Having adopted a formal Bayesian approach to the probabilistic analysis of the model allows the intuitive interpretation often afforded to acceptability curves as showing the probability that the intervention is cost-effective. Note that the curves all sum to one on the vertical axis (this clearly must be the case since only one strategy is chosen for each value of the ceiling ratio and for each replication of the model). It is immediately apparent from both Figures 3 and 4 that strategy D is always dominated. The acceptability curves in Figure 4 clearly show that initial concern that strategy F might form part of the frontier was unwarranted. In fact, the conditions necessary for strategy F to be considered the most cost-effective option rarely arose in the simulations.

While the sort of presentation of the choice between mutually exclusive treatments in the face of many options is a natural extension of the use of cost-effectiveness acceptability curves in the two treatment case, the issue arises of how exactly decision-makers are expected to use this information to choose between the remaining strategies that form part of the frontier. One approach (as illustrated in Figure 5) would be to say that for any given value of the shadow price, the optimal decision would be to choose the strategy that is most likely to be cost-effective. But of course, this decision rule completely ignores the uncertainty that has just been modelled and gives the exact same treatment recommendations as the base line estimates in Figure 2.

The conventional approach to statistical decision-making is based on the adoption of a 5% type I error rate. We might therefore adopt a decision-rule that a more effective and more expensive treatment strategy should only replace the currently provided treatment if it can be shown to be significantly more cost-effective. This approach to decision-making is illustrated in Figure 6 and gives markedly different cut-off points for decision-maker’s ceiling ratios for the different strategies to be considered cost-effective. For example, the most effective therapy, step-down maintenance PPI, would only be considered the appropriate treatment option if decision-makers had an underlying willingness to pay of $480 per week free of GERD symptoms.

Note that under a 5% significance decision rule, neither strategy A or E would be considered significantly cost-effective to be a clear decision choice. However, it is clear that the basic strategy C is not cost-effective (at the 5% error rate) for a shadow price greater than $26 per week free of GERD symptoms. Recall that strategy A involves healing with PPI without maintenance and strategy E involves healing with PPI then maintenance using H$_2$RAs. The choice
between these strategies is between no maintenance and maintenance with H$_2$RAs following healing. If the shadow price of a week free of GERD symptoms is between $26 and $480 per week free of GERD symptoms then we know that strategies B, C, D and F are not cost-effective. We cannot distinguish strategies A & E at conventional significance levels so for purposes of decision making it should be clear that the strategy to heal with PPI should be adopted, but that it is unlikely to be important whether, subsequent to healing, no maintenance or maintenance with H$_2$RAs is undertaken. This is illustrated in Figure 6 by combining the two strategies in one acceptability curve.

It is important to recognise, however, the arbitrary nature of the conventional decision rule. Consider if instead of placing the ‘burden of proof’ for cost-effectiveness on more expensive and more effective strategies, it is the cheaper but less effective strategies that would only be used if they were shown to be significantly cost-effective. Although not shown in Figure 6, it should be clear from the above exposition that this change in the burden of proof would result in a new set of threshold values such that strategy B would be the treatment of choice unless the shadow price were below $170 per week free of GERD symptoms; between $17.50 and $170 per week free of GORD symptoms either strategy A or E would be considered cost-effective; and below a shadow price of $17.50, either A or C would be the strategy of choice.

Of course the arbitrary nature of such decision-making under uncertainty emphasises the inadequacies of such a simple decision rule – Claxton has argued that significance testing of this sort is irrelevant (Claxton, 1999). Instead he suggests that decision-making should be concerned fundamentally with expected values. That is not to say that the decisions should be made on the basis of the baseline point estimates as presented in Figure 2 without reference to uncertainty in obtaining those estimates. Rather, that the expected returns to obtaining further information should be assessed in order to determine whether it is worth commissioning more research to obtain improved estimates of the decision parameters. Such an approach would require estimates of: the loss function associated with incorrect decision-making; the size of the population relevant to the decision; the lifetime of the technologies associated with each management strategy; and the returns to sampling. Each of these things is itself subject to a great deal of uncertainty and the methods for incorporating all this information into one overall analysis are currently under development. We have not attempted such an analysis in this paper, although we recognise that this is a possible extension to the work presented here.

In summary, probabilistic modelling of deterministic models is a practical solution to the problems of conventional sensitivity analysis. Adopting a formal Bayesian approach encour-
ages analysts and users to think carefully about the state of evidence relating to the parameters of the model. Single parameter specifications are straightforward to apply in a Bayesian framework and provide a simple way to update parameter distributions as new data becomes available. The use of acceptability curves to present information on the probability of multiple treatment options is a natural extension of the two alternative case usually presented in the literature. Much current research interest is focused on expected value of information methods. It is clear that such methods will have to be predicated on a well specified probabilistic model.
Probabilistic Analysis of Cost-Effectiveness Models: Choosing Between Treatment Strategies for Gastro-Esophageal Reflux Disease
Reference List


Table 1  Step-up and switching algorithms conditional upon healing failure or recurrence

<table>
<thead>
<tr>
<th>Strategy A: Intermittent PPI</th>
<th>Strategy B: Maintenance PPI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Healing</strong></td>
<td><strong>Maintenance</strong></td>
</tr>
<tr>
<td>PPI</td>
<td>No therapy</td>
</tr>
<tr>
<td>↓ unhealed</td>
<td></td>
</tr>
<tr>
<td>DD PPI</td>
<td>PPI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy C: Maintenance H₂RA</th>
<th>Strategy D: Step-Down Maintenance PA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Healing</strong></td>
<td><strong>Maintenance</strong></td>
</tr>
<tr>
<td>H₂RA</td>
<td>H₂RA</td>
</tr>
<tr>
<td>↓ unhealed</td>
<td></td>
</tr>
<tr>
<td>PPI</td>
<td>H₂RA</td>
</tr>
<tr>
<td>↓ unhealed</td>
<td></td>
</tr>
<tr>
<td>DD PPI</td>
<td>PPI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy E: Step-Down Maintenance H₂RA</th>
<th>Strategy F: Step-Down Maintenance PPI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Healing</strong></td>
<td><strong>Maintenance</strong></td>
</tr>
<tr>
<td>PPI</td>
<td>H₂RA</td>
</tr>
<tr>
<td>↓ unhealed</td>
<td></td>
</tr>
<tr>
<td>DD PPI</td>
<td>PPI</td>
</tr>
</tbody>
</table>

PPI - Proton Pump Inhibitor (e.g., omeprazole 20mg OD)  
Pump inhibitor (e.g., omeprazole 40mg OD)  
LD PPI - Low Dose Proton Pump Inhibitor (e.g., omeprazole 10mg OD)  
Antagonists (e.g., ranitidine 150mg BID)  
DD H₂RA - Double Dose H₂ Receptor Antagonists (e.g., ranitidine 300mg BID)  
cisapride 10mg QID)  
LD PA - Low Dose Prokinetic Agent (e.g., cisapride 10mg BID)  
DD PPI - Double Dose Proton  
H₂RA - H₂ Receptor  
PA - Prokinetic Agent (e.g.,
### Table 2 Parameters in the model

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Healing hazards</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hPA</td>
<td>0.045</td>
<td>hazard for healing on PA</td>
</tr>
<tr>
<td>hH2</td>
<td>0.068</td>
<td>hazard for healing on H2RAs</td>
</tr>
<tr>
<td>hDDH2</td>
<td>0.091</td>
<td>hazard for healing on double dose H2RAs</td>
</tr>
<tr>
<td>hPPI</td>
<td>0.215</td>
<td>hazard for healing on PPIs</td>
</tr>
<tr>
<td>hDDPPI</td>
<td>0.209</td>
<td>hazard for healing on double dose PPIs</td>
</tr>
<tr>
<td><strong>Corresponding transition probabilities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pPA</td>
<td>0.42</td>
<td>probability of healing on PA (at 12 weeks)</td>
</tr>
<tr>
<td>pH2</td>
<td>0.42</td>
<td>probability of healing on H2RAs (at 8 weeks)</td>
</tr>
<tr>
<td>pDDH2</td>
<td>0.52</td>
<td>probability of healing on double dose H2RAs (at 8 weeks)</td>
</tr>
<tr>
<td>pPPI</td>
<td>0.82</td>
<td>probability of healing on PPIs (at 8 weeks)</td>
</tr>
<tr>
<td>pDDPPI</td>
<td>0.81</td>
<td>probability of healing on double dose PPIs (at 8 weeks)</td>
</tr>
<tr>
<td><strong>Recurrence variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p06PL</td>
<td>0.67</td>
<td>probability of recurrence on placebo (0-6months)</td>
</tr>
<tr>
<td>p06PA</td>
<td>0.23</td>
<td>probability of recurrence on PA (0-6months)</td>
</tr>
<tr>
<td>p06H2</td>
<td>0.38</td>
<td>probability of recurrence on H2RAs (0-6months)</td>
</tr>
<tr>
<td>p06PPI</td>
<td>0.12</td>
<td>probability of recurrence on PPIs (0-6months)</td>
</tr>
<tr>
<td>p06LDPPI</td>
<td>0.25</td>
<td>probability of recurrence on low dose PPIs (0-6months)</td>
</tr>
<tr>
<td>p06SU</td>
<td>0.12</td>
<td>probability of recurrence after surgery (0-6months)</td>
</tr>
<tr>
<td>p612PL</td>
<td>0.23</td>
<td>probability of recurrence on placebo (6-12months)</td>
</tr>
<tr>
<td>p612PA</td>
<td>0.13</td>
<td>probability of recurrence on PA (6-12months)</td>
</tr>
<tr>
<td>p612H2</td>
<td>0.18</td>
<td>probability of recurrence on H2RAs (6-12months)</td>
</tr>
<tr>
<td>p612PPI</td>
<td>0.08</td>
<td>probability of recurrence on PPIs (6-12months)</td>
</tr>
<tr>
<td>p612LDPPI</td>
<td>0.13</td>
<td>probability of recurrence on low dose PPIs (6-12months)</td>
</tr>
<tr>
<td><strong>Symptom-week variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWPA</td>
<td>2.280</td>
<td>symptom weeks on PA</td>
</tr>
<tr>
<td>SWH2</td>
<td>1.521</td>
<td>symptom weeks on H2RAs</td>
</tr>
<tr>
<td>SWDDH2</td>
<td>1.816</td>
<td>symptom weeks on double dose of H2RAs</td>
</tr>
<tr>
<td>SWPPI</td>
<td>2.386</td>
<td>symptom weeks on PPIs</td>
</tr>
<tr>
<td>SWDDPPI</td>
<td>2.383</td>
<td>symptom weeks on double dose of PPIs</td>
</tr>
<tr>
<td>SWSU</td>
<td>0.539</td>
<td>symptom weeks after surgery</td>
</tr>
<tr>
<td>Name</td>
<td>Value</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Fee</td>
<td>4.11</td>
<td>Dispensing fee</td>
</tr>
<tr>
<td>H2RA</td>
<td>0.44</td>
<td>150mg H2RA</td>
</tr>
<tr>
<td>CIS</td>
<td>0.61</td>
<td>10mg PA</td>
</tr>
<tr>
<td>PPI</td>
<td>2.42</td>
<td>20mg PPI</td>
</tr>
<tr>
<td>LDPII</td>
<td>1.93</td>
<td>10mg PPI</td>
</tr>
<tr>
<td>GPGA</td>
<td>48.20</td>
<td>GP general assessment</td>
</tr>
<tr>
<td>GPRA</td>
<td>28.10</td>
<td>GP re-assessment</td>
</tr>
<tr>
<td>GPMA</td>
<td>16.25</td>
<td>GP minor assessment</td>
</tr>
<tr>
<td>GERA</td>
<td>38.65</td>
<td>gastroenterologist re-assessment</td>
</tr>
<tr>
<td>GEPA</td>
<td>23.10</td>
<td>gastroenterologist partial assessment</td>
</tr>
<tr>
<td>UGIE</td>
<td>118.22</td>
<td>upper GI endoscopy</td>
</tr>
<tr>
<td>UGIS</td>
<td>141.43</td>
<td>upper GI series</td>
</tr>
<tr>
<td>CST</td>
<td>84.81</td>
<td>cardiac stress test</td>
</tr>
<tr>
<td>ECG</td>
<td>42.77</td>
<td>ECG</td>
</tr>
<tr>
<td>BS</td>
<td>135.10</td>
<td>barium swallow</td>
</tr>
<tr>
<td>NF</td>
<td>2,462.60</td>
<td>laparoscopic fundal plication</td>
</tr>
</tbody>
</table>

**Resource use variables**

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>nGPR1</td>
<td>2.5</td>
<td>Visits to GP (first recurrence)</td>
</tr>
<tr>
<td>nGER1</td>
<td>1.5</td>
<td>Visits to gastroenterologist (first recurrence)</td>
</tr>
<tr>
<td>nBSR1</td>
<td>0.1</td>
<td>Percentage getting a barium swallow (first recurrence)</td>
</tr>
<tr>
<td>nCSTR1</td>
<td>0.0</td>
<td>Percentage getting cardiac stress test (first recurrence)</td>
</tr>
<tr>
<td>nECGR1</td>
<td>0.0</td>
<td>Percentage getting ECG (first recurrence)</td>
</tr>
<tr>
<td>nUGIER1</td>
<td>0.1</td>
<td>Percentage getting upper GI endoscopy (first recurrence)</td>
</tr>
<tr>
<td>nUGISR1</td>
<td>0.1</td>
<td>Percentage getting upper GI series (first recurrence)</td>
</tr>
<tr>
<td>nGPR2</td>
<td>2.5</td>
<td>Visits to GP (second recurrence)</td>
</tr>
<tr>
<td>nGER2</td>
<td>1.5</td>
<td>Visits to gastroenterologist (second recurrence)</td>
</tr>
<tr>
<td>nBSR2</td>
<td>0.0</td>
<td>Percentage getting a barium swallow (second recurrence)</td>
</tr>
<tr>
<td>nCSTR2</td>
<td>0.0</td>
<td>Percentage getting cardiac stress test (second recurrence)</td>
</tr>
<tr>
<td>nECGR2</td>
<td>0.0</td>
<td>Percentage getting ECG (second recurrence)</td>
</tr>
<tr>
<td>nUGIER2</td>
<td>0.6</td>
<td>Percentage getting upper GI endoscopy (second recurrence)</td>
</tr>
<tr>
<td>nUGISR2</td>
<td>0.1</td>
<td>Percentage getting upper GI series (second recurrence)</td>
</tr>
</tbody>
</table>
Figure 1 Decision tree for the management of erosive esophagitis

Endoscopically Proven Erosive Esophagitis

- **Strategy A: PPI**
  - Healed → No Prescription Therapy
  - Not Healed → Step up therapy (Table 1)

- **Strategy B: PPI**
  - Healed → Maintenance PPI
  - Not Healed → Step up therapy (Table 1)

- **Strategy C: H₂ RA**
  - Healed → Maintenance H₂ RA
  - Not Healed → Step up therapy (Table 1)

- **Strategy D: PA**
  - Healed → Maintenance Low Dose PA
  - Not Healed → Step up therapy (Table 1)

- **Strategy E: PPI**
  - Healed → Maintenance H₂ RA
  - Not Healed → Step up therapy (Table 1)

- **Strategy F: PPI**
  - Healed → Maintenance Low Dose PPI
  - Not Healed → Step up therapy (Table 1)

0-6 MONTHS

- Recurrence (step up therapy -- see Table 1)
  - No Recurrence → No Recurrence

6-12 MONTHS

- Recurrence (step up therapy -- see Table 1)
  - No Recurrence → No Recurrence
Figure 2 Baseline cost-effectiveness results on the cost-effectiveness plane showing the ‘efficient frontier’

- A: Intermittent PPI
- B: Maintenance PPI
- C: Maintenance H₂RA
- D: Step-down maintenance PA
- E: Step-down maintenance H₂RA
- F: Step-down maintenance PPI

Weeks free of GORD

Strategy cost

$10/GFW $36/GFW $284/GFW
Figure 3 Results of 10,000 Monte Carlo simulation evaluations of the GERD model presented on the cost-effectiveness plane

A: Intermittent PPI
B: Maintenance PPI
C: Maintenance H₂RA
D: Step-down maintenance PA
E: Step-down maintenance H₂RA
F: Step-down maintenance PPI

Weeks free of GERD

Strategy cost

$1,200
$1,100
$1,000
$900
$800
$700
$600

38.00 39.00 40.00 41.00 42.00 43.00 44.00 45.00 46.00 47.00 48.00
Figure 4 Acceptability curves for the choice of treatment strategy (a log scale is employed to better illustrate the low values)
Figure 5 Decision-making using management strategy 'most likely' to be cost-effective.
Figure 6 Decision making assuming more expensive strategies have to be shown to be significantly more cost-effective than the current strategy (at conventional 5% level)
Probabilistic Analysis of Cost-Effectiveness Models: Choosing Between Treatment Strategies for Gastro-Esophogeal Reflux Disease