International Journal of Technology Assessment in Health Care, **17:1** (2001), 83–97.

BAYESIAN COST-EFFECTIVENESS ANALYSIS

An Example Using the GUSTO Trial

Dennis G. Fryback
University of Wisconsin-Madison

James O. Chinnis, Jr.
Jacob W. Ulvila
Decision Science Associates, Inc.

Abstract

A desirable element of cost-effectiveness analysis (CEA) modeling is a systematic way to relate uncertainty about input parameters to uncertainty in the computational results of the CEA model. Use of Bayesian statistical estimation and Monte Carlo simulation provides a natural way to compute a posterior probability distribution for each CEA result. We demonstrate this approach by reanalyzing a previously published CEA evaluating the incremental cost-effectiveness of tissue plasminogen activator compared to streptokinase for thrombolysis in acute myocardial infarction patients using data from the GUSTO trial and other auxiliary data sources. We illustrate Bayesian estimation for proportions, mean costs, and mean quality-of-life weights. The computations are performed using the Bayesian analysis software WinBUGS, distributed by the MRC Biostatistics Unit, Cambridge, England.

Keywords: Bayes theorem, Cost-effectiveness analysis, Sensitivity analysis

A Bayesian formulation of a cost-effectiveness analysis can lead in a straightforward way to a probability distribution expressing uncertainty about any particular numerical result of the analysis. In this paper we discuss briefly the Bayesian conceptual formulation for a cost-effectiveness analysis (CEA), and motivate why one might want to use such a formulation. Next, we illustrate such a computation, using as an example thrombolytic therapy for acute myocardial infarction with tissue plasminogen activator (t-PA) compared to streptokinase. Finally, we summarize with thoughts about Bayesian analysis of cost-effectiveness data in general.

THE BAYESIAN FORMULATION

An incremental CEA is a complex numerical model (14). Quantities such as the incremental cost, ΔC, or the incremental effectiveness, ΔE, of one medical intervention over its

The authors thank Karl Claxton, Andrew Briggs, Milton Weinstein, Natasha Stout, and Marjorie Rosenberg for valuable comments on earlier drafts of this work. This work was presented in part at the Annual Scientific Meeting of the Society for Medical Decision Making, Boston, MA, October 1998. Funded in part by a contract under The Bayesian Initiative, MEDTAP International Inc., Bethesda, MD.

comparator are functions of many input parameters, as are additional functions of these two quantities. One common such function in CEA is the incremental cost-effectiveness ratio, $ICER = \Delta C/\Delta E$. This can be written as $ICER = f(\theta_1, \theta_2, \ldots, \theta_k; \tau_1, \tau_2, \ldots, \tau_m)$, where the θ_i are stochastic parameters and the τ_j are design (fixed) parameters.

Examples of the latter parameters include the screening interval or age at initial screen (when analyzing a screening program) or the dosage of a drug (when analyzing a pharmacologic intervention). The θ_i are parameters that include, for example, compliance probabilities, test sensitivity and specificity, complication rates, survival rates or survival times, and average quality of life of patients in various outcome groups, and average costs for various health resources expended in the treatment. In practice these usually are estimated from primary data, such as those collected in a prospective randomized controlled study of the intervention, by secondary use of data from published or archived sources, or they are estimated by experts. Most importantly, each of these inputs to the CEA model can be considered a random variable. CEA uses the best point estimate (i.e., best single numerical value, often the mean of some appropriate sample of data) for each of these inputs to compute the base case cost-effectiveness results. In traditional CEA, uncertainty about the true values of the parameters is acknowledged by using sensitivity analyses to explore the effect of variation in these estimates on the computed outputs of the CEA (6).

The Bayesian formulation is concerned with derivation of a posterior probability distribution for the vector θ as a function of some currently observed data:

$$p(\theta \mid d) = \frac{p(d \mid \theta) \cdot p(\theta)}{p(d)},$$

where $p(\theta \mid d)$ is the joint density of the θ_i conditioned on the observed data, d. Mathematically, $p(d)$, the marginal likelihood of the observed data, plays the role of a normalizing constant, and modern methods for evaluating the expression ignore it. The two important pieces of the expression are in the numerator of the right-hand side. Prior to using the data at hand, we express our uncertainty about the true value of θ as a probability density function, $p(\theta)$. While $p(\theta)$ does contain information from previously observed data and opinion about likely values of θ, it contains no information about the currently observed data. These data are accounted for through the data likelihood function, $p(d \mid \theta)$, which is the joint probability density of the observed data given knowledge of the true parameter values. This expression, known as Bayes' theorem, converts prior opinion about θ into posterior opinion (we hope with more precision than the prior) by taking into account the newly observed data. From the resulting density, $p(\theta \mid d)$, we can calculate $p(\mathrm{f}(\theta, \tau) \mid d)$ as an induced posterior probability density on $\mathrm{f}(\theta, \tau)$, the ICER. In principle, any numerical result calculated in a CEA will have an associated probability distribution that can be expressed this way, and later we will replace the ICER with an expression of net benefit. Uncertainty about each of the inputs is carried through to the final expression of uncertainty about the computational outputs. In practice, putting this formulation to work will involve many assumptions and compromises. We will discuss some of these later.

Why do we want to derive a distribution over outputs of a cost-effectiveness analysis? Elsewhere Claxton (4) has argued that inference about the cost-effectiveness ratio (i.e., deriving a probability distribution for it) is irrelevant. He notes that decisions will be made on the mean of this distribution (the point estimate of the ratio, which can be calculated without the full distribution), so the whole distribution is not needed. However, there remain two uses for the distribution.

The less quantitative of these uses is merely to characterize our uncertainty about analytic result. Consumers of CEA results worry about how "soft" the results might be, since many quantities going into the CEA calculations are not known with certitude. The

effect of this on the robustness of the results has been explored using deterministic sensitivity analyses, and more recently, various forms of stochastic sensitivity analyses (for discussions of these approaches see 1;6,248-255;15;18).

The more formal quantitative use of posterior distributions for the CEA output is in value of information analysis. As Claxton (4) discusses, these distributions can be used to compute the Bayes risk of associated decisions, and the decision maker can evaluate whether collecting more information about input parameters is worthwhile in the face of the expected gains and losses of the decisions and the costs of additional information about the uncertain quantities in the analysis.

RELATIONSHIP BETWEEN BAYESIAN AND PROBABILISTIC SENSITIVITY ANALYSES IN CEA

A Bayesian CEA is a form of probabilistic sensitivity analysis. To perform a probabilistic sensitivity analysis of a CEA result, one first specifies probability distributions describing the uncertainty about the critical numerical inputs to the model. These may be univariate or multivariate distributions as appropriate to the problem and knowledge about the inputs. Next, the distributions are sampled once, and the sampled parameter values are used to compute the CEA results. Sampling and calculation are repeated hundreds or thousands of times, and the set of computed CEA results can be used to derive an empirical distribution for the CEA output (5). In practice, this is the process of a Bayesian analysis as well, and the posterior distributions of the Bayesian analysis form a probabilistic sensitivity analysis.

Beyond this, a Bayesian approach to CEA might be better described in light of a specific example. We present an example next, and return to discussion about distinguishing characteristics of a Bayesian analysis in the context of this example.

BAYESIAN CEA OF TISSUE PLASMINOGEN ACTIVATOR VERSUS STREPTOKINASE IN ACUTE MYOCARDIAL INFARCTION: AN EXAMPLE

We have picked this problem solely to illustrate Bayesian CEA and not necessarily to be informative on the substantive problem. We found the problem attractive for two reasons: a) a non-Bayesian CEA has been published (16); and b) a Bayesian analysis has been published examining the primary results this CEA was based on (3). We have relied on the published reports without access to the original primary data from the trial. Our calculations are presented for pedagogic use in presenting methods. However, we believe our results are intriguing and may stimulate others to pursue further analyses using primary data to confirm and extend our calculations.

Background

The early 1990s saw several trials of tissue plasminogen activator (t-PA), a newly engineered agent, compared with an existing agent, streptokinase (SK), administered to patients experiencing acute myocardial ischemia or infarction (MI) to dissolve or break clots occluding coronary arteries. The newer agent was priced an order of magnitude more than the older one ($2,750 versus $320 per dose) and, if t-PA were shown to reduce mortality in acute MI compared with SK, there were considerable cost implications. Two large European trials showed no survival advantage to t-PA and, indeed, perhaps a slight reduction in short-term survival compared with SK (3). A later, well-conducted international trial, Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries (GUSTO), used an administration protocol differing from the earlier trials. In GUSTO, accelerated administration of t-PA to 10,344 patients showed a reduction in 30-day mortality

of 1% among patients receiving t-PA, but a 0.1% increase in strokes among the surviving patients compared with 20,173 patients treated with SK (11). (Both arms received heparin as well as the thrombolytic drugs being investigated.)

The CEA published from GUSTO data used the GUSTO point estimates for survival and stroke in the two groups and concluded that the administration of t-PA versus SK resulted in an incremental cost-effectiveness ratio of $36,402 per quality-adjusted life-year (QALY) saved (16). A concurrent Bayesian analysis of the GUSTO result, giving full weight to the prior trial results—a controversial decision—concluded that the trial evidence aggregated over all trials slightly favored SK over t-PA (3).

In our example analysis, we reformulate the CEA in Mark et al. (16) as a Bayesian CEA.

The CEA model

Our model has 12 critical parameters—the θ_i noted earlier. Inferences about these parameters are summarized in the following sections. The parameters are divided into three categories: a) proportions; b) costs; and c) quality-of-life weights. We will discuss inferences on each of these types to obtain distributions, $p(\theta_i \mid d)$, to describe our input uncertainty about each. At the outset we note that, without access to primary data, we will necessarily have to assume these are independent parameters, as did the earlier published CEA.

Inferences About Proportions. There are four proportions used in the model. Two are the proportion of patients administered accelerated t-PA who survive 30 days and the proportion of survivors who experience a nonfatal disabling stroke during those 30 days (a possible complication of the thrombolytic therapy). The GUSTO-derived point estimates for these proportions are 93.7% (or 9,691/10,343), and 0.64%, or 62 of 9,691, respectively, according to Table 1 in reference 3. Our reading of the original GUSTO report is that these are 9,692/10,344 and 62/9,692, respectively, and we shall use these fractions for the following analyses. Corresponding data were observed for patients in the two SK arms of the trial. Combined across these two arms is a total of 18,700 of 20,173 patients, or 92.7%, surviving 30 days, and 101 of $18{,}700 = 0.54\%$ of survivors suffering nonfatal disabling stroke in the first 30 days.

Brophy and Joseph (3) note that these point estimates do not incorporate knowledge already gained in the previous trials and demonstrate a Bayesian analysis incorporating the earlier results with varying weights. By including the earlier results, the aggregate estimate for 30-day survival with t-PA is lower, dropping to equal or below the SK survival estimate. Their analysis has been criticized because the protocols for administering t-PA in those earlier trials were different from the accelerated protocol in GUSTO.

In a Bayesian analysis, uncertainty about an unknown proportion, θ, may be conveniently described by a beta distribution. The beta density is a function of two parameters, a and b, and is defined on the interval between 0 and 1:

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1},$$

where $0 \leq \theta \leq 1$, and $a > 0$, $b > 0$.

In this equation, $\Gamma(\cdot)$ stands for the gamma function, a particular integral used commonly in statistics. Rather than write this function repeatedly, we will denote the beta density as beta(a, b). The parameters a and b can be thought of as the number of "successes" and number of "failures" in a sequence of observed data from Bernoulli trials, with the trials being mutually independent and the probability of success in each being θ. The mean of the beta distribution is $a/(a+b)$, and this is the usual point estimate for θ. The larger the sum $a + b$, the smaller the variance around this mean. Beta(1, 1) is the familiar uniform

density on the interval [0, 1], with a mean of 0.5. When a and b are both less than 1, the beta density has a "U" shape. When a and b are both larger than 1, the density has a single peak; when a and b are very large, the beta density looks much like a normal density over a very short interval within the [0, 1] range. All major computer spreadsheet programs have statistical functions built in for the beta distribution. (Caution is needed to verify the form in which the beta distribution is parameterized since some programs may use the alternate with parameters r and m, where $r = a$ and $m = a + b$.)

Data we would usually collect to estimate a proportion can be represented as a sequence of logical zeroes and ones; for example, in the GUSTO data, a patient who dies in the first 30 days can be denoted by zero and a patient who survives denoted by one, with the trial results represented by a string of 10,344 zeroes and ones for the t-PA arm. The likelihood that any given case is a one is the unknown proportion θ. The data then are Bernoulli observations with probability θ. Collectively, if we observe N cases among which there are k ones and m zeroes, this is one observation of a binomial variable with parameters k and N (where $N = k + m$). If our uncertainty about the value of θ before the trial results are known is described by beta(a, b), then applying Bayes theorem will show that our uncertainty about θ at the end of the trial is described by beta($a + k$, $b + m$), which is termed the posterior density for θ given the observed data, and often denoted $p(\theta \mid d)$. Because the prior and the posterior distributions are the same type, the beta distribution and the Bernoulli distribution (or its alternate form, the binomial distribution) are known as *conjugate* distributions. In the language of Bayesian statistics, a beta prior, updated with Bernoulli data, gives a beta posterior. If our best point estimate of θ is, say, 30%, but we still have considerable uncertainty around this point estimate, our uncertainty might be described by a beta(3, 7) density. If we are more certain that 30% is the true value, we would use a beta distribution with smaller variance, for example a beta(9, 21) might do (each parameter multiplied by 3). If we are quite sure, then beta(30, 70) or beta(300, 700) or beta(3,000, 7,000) or beta(9,000, 21,000) might be the best description. Table 1 shows this succession of beta density functions along with the lower and upper bounds for the central 95% of each distribution (in Bayesian statistics, this is known as the 95% credible interval). Note that updating the Beta prior is just a function of adding to the prior parameters the observed numbers of successes and failures in the data.

Our Bayesian analysis of the four relevant proportions from the GUSTO trial will use the beta-Bernoulli (or beta-binomial) conjugate relationship. In light of the GUSTO data, our posterior uncertainty about the survival proportion in patients treated with SK after the GUSTO trial can be represented as beta($a + 18{,}700$, $b + 1{,}473$), where beta(a, b) is the prior and 18,700 and 1,473 are the specific results for 30-day survival and 30-day deaths, respectively, in the SK arms of the trial. We could assume an noninformative prior distribution (this prior has essentially no impact on the posterior) where a and b are very

Table 1. Beta Density Functions with Increasing Certainty About the Point Estimate of 30%

		95% central interval bounds	
Beta density	Mean	Lower	Upper
Beta (0.3, 0.7)	30%	0.0008%	97.87%
Beta (3, 7)	30%	7.49%	60.01%
Beta (9, 21)	30%	15.28%	47.24%
Beta (30, 70)	30%	21.47%	39.29%
Beta (300, 700)	30%	27.20%	32.88%
Beta (3,000, 7,000)	30%	29.11%	30.90%
Beta (9,000, 21,000)	30%	29.48%	30.52%

close to zero, so the posterior will be closely approximated by beta(18,700, 1,473) which has a mean equal to the observed point estimate from the GUSTO trial, 18,700 of 20,173. But the survival of patients given streptokinase in the two previous trials (as summarized by Brophy and Joseph) was 9,467 of 10,396 and 12,325 of 13,780. If these trials are considered to be equivalent to the GUSTO SK treatment arms, the combined result, 21,792/24,176 may well form a prior estimate of the survival proportion for patients given SK, i.e., our prior is beta(21,792, 2,384), with a mean of 90.1%.

These are the two extremes for incorporating prior observed data—either use an entirely noninformative prior so no previous data are included, or use a prior that gives all previous data full weight. We will elect an intermediate position here, neither giving full weight nor giving no weight to the prior trials. Because the SK arms in the earlier trials were relevant but not entirely so, even if just from passage of time with improvement in survival, we will keep the same mean, but widen the credible interval by cutting the sum of the parameters to one-third of the nominal sample size, beta(7,264, 794.7). This allows the data from GUSTO to have about 70% of the weight in determining the posterior, even though the total number of patients in GUSTO is only about 45% of the aggregate SK patients in the other trials. By similar reasoning, we use a prior of beta(43.7, 8,015) on the proportion of nonfatal disabling strokes in the first 30 days among surviving patients treated with SK.

What shall we use for the prior distributions on these two proportions (30-day survival and 30-day stroke) in the t-PA arm of GUSTO? The protocol for administration of t-PA in GUSTO was changed from that in the earlier trials and the new accelerated protocol was expected, based on a series of limited trials using patency of coronary arteries at angiography as an endpoint (e.g., see reference 17), to improve patient results over the prior trials with nonaccelerated administration of t-PA. Still, the prior trials found 30-day survival with t-PA to be high, an aggregate result of 21,707 of 24,118, or 90.0%, almost exactly the same as SK. The GUSTO trial was conducted to test the hypothesis that accelerated administration of t-PA might better this percentage. We will use a prior with this mean, but much more variance to represent our broader range of uncertainty. The beta (45, 5) has a mean of 90%, with a 95% credible interval of (80.4%, 96.0%). The density is peaked toward the upper end of the credible interval (the median is 90.5%), to show that the anticipated result is in the neighborhood of 90%, but more likely over 90% than under. With this prior, the posterior density after the GUSTO data are included is beta $(45 + 9{,}692,\ 5 + 652) =$ beta (9,737, 657). The mean of the posterior is 93.68%, with a 95% credible interval of (93.2%, 94.1%). Similar reasoning is applied to the proportion who have nonfatal disabling stroke in the first 30 days. Table 2 summarizes the input distributions for the four proportions in the CEA.

We have gone to some detail in describing the Bayesian estimation of the four proportions. Bayesian analysis of data concerning proportions is well known and made easy by the conjugacy of the beta and binomial distributions. For the two other types of inputs to

Table 2. Prior and Posterior Beta Densities Used to Describe Uncertainty About the Four Proportions to be Used in the Cost-effectiveness Analysis

		Posterior		
Estimated proportion	Prior density	Density	Mean (%)	95% credible interval
t-PA 30-day survival	Beta (45, 5)	Beta (9,737, 657)	93.7	93.2, 94.1
SK 30-day survival	Beta (7,264, 794.7)	Beta (25,964, 2,267.7)	92.0	91.6, 92.3
t-PA stroke	Beta (1.7, 215.4)	Beta (63.7, 9,845.4)	0.64	0.50, 0.81
SK stroke	Beta (131, 21,661)	Beta (232, 40,260)	0.57	0.50, 0.65

https://doi.org/10.1017/S0266462301104083 Published online by Cambridge University Press

this CEA, we do not have the benefit of conjugate distributions and must use a different method to derive posterior distributions given the data and prior uncertainties.

Inferences About Quality-of-life Weights. For cost-utility analysis time spent in various health states is weighted by the health-related quality of life (HRQL) deemed to characterize those health states. HRQL is indicated by a number anchored by 1.0, denoting perfect health, and 0.0, denoting death. For purposes here, we will ignore states worse than death and consider HRQL to be a number between 0 and 1. In this fashion it is similar to a proportion, and in fact HRQL weights that are elicited by the method of standard gambles or by time trade-offs are proportions. But estimates of each are collected directly and not by observation of 0 and 1 data. In the GUSTO trial, a sample of patients in each arm were queried at the end of the trial about their post-MI HRQL using the time tradeoff technique, and the result was a mean weight of 0.90 in both arms (16). There are other data in the literature directly relevant to this HRQL weight. The Beaver Dam Health Outcomes Study (9) using time tradeoffs found an average weight of 0.73 among community-dwelling adults, aged 45 and older, who reported they had had an MI in the past. Tsevat et al. (22) reported a mean of 0.88 among MI survivors. We used a beta(16, 4) to represent uncertainty about average HRQL of MI survivors in years following the MI. This is a posterior distribution directly estimated by us to describe the disparate results in the literature since we don't know the data-generating process by which to model these reports. (There are other ways to approach specifying this distribution on post-MI HRQL. For example, with access to primary data, we might pool the data and process the data as described below for stroke HRQL. Given the limited use of this example for pedagogic purposes, we have elected not to go beyond the direct specification approach.) The mean of this directly specified posterior distribution is 0.80. Its median is 0.810, and the 95% credible interval ranges from 0.604 to 0.940. This shows a considerable range of uncertainty and covers all three estimates. It does represent a lower mean than observed just among patients in the GUSTO trial, reflecting other data in the literature to inform this estimate. The effect of this will be to devalue slightly the life-years saved in our calculations as compared with the GUSTO CEA.

The GUSTO CEA did not directly consider the HRQL associated with disabling stroke; a sensitivity analysis assigned this condition a weight of zero, equivalent to death. Instead, we used data on HRQL weight assigned to disabling stroke by 1,176 people with stroke or at high risk of stroke who were interviewed by investigators of the Stroke Patient Outcome Research Team at Duke University (personal communication, Dr. Greg Samsa, 1998). A considerable fraction of these people assigned a weight of zero to this condition; the remainder spread across the continuum. Figure 1, upper panel, shows the empirical distribution of HRQL weights. The lower panel in Figure 1 shows the estimated posterior density for the mean HRQL weight given these data and assuming an informationless prior. The posterior density is estimated using Markov Chain Monte Carlo computations carried out by WinBUGS version 1.2 (distributed by The BUGS Project group, MRC Biostatistics Unit, Cambridge, U.K.). For a brief overview of WinBUGS syntax, see Fryback DG, Stout NK, Rosenberg MA, "An Elementary Introduction to Bayesian Computing Using WinBUGS" in this Special Section.

Parameters for a normal distribution were fitted to the simulated posterior kernel density using the method of moments; a normal distribution with a mean of 0.234 and a precision (the reciprocal of the variance) of 14,550.9 was then used to represent uncertainty about mean HRQL with disabling stroke in further computations.

Inferences About Costs. The third type of variable for the CEA computations is mean cost. In this analysis we considered six costs: a) the pharmaceutical agent cost for t-PA; b) the pharmaceutical agent cost for SK; c) medical costs in the first year for patients in the t-PA treatment group; d) medical costs in the first year for patients in the SK treatment

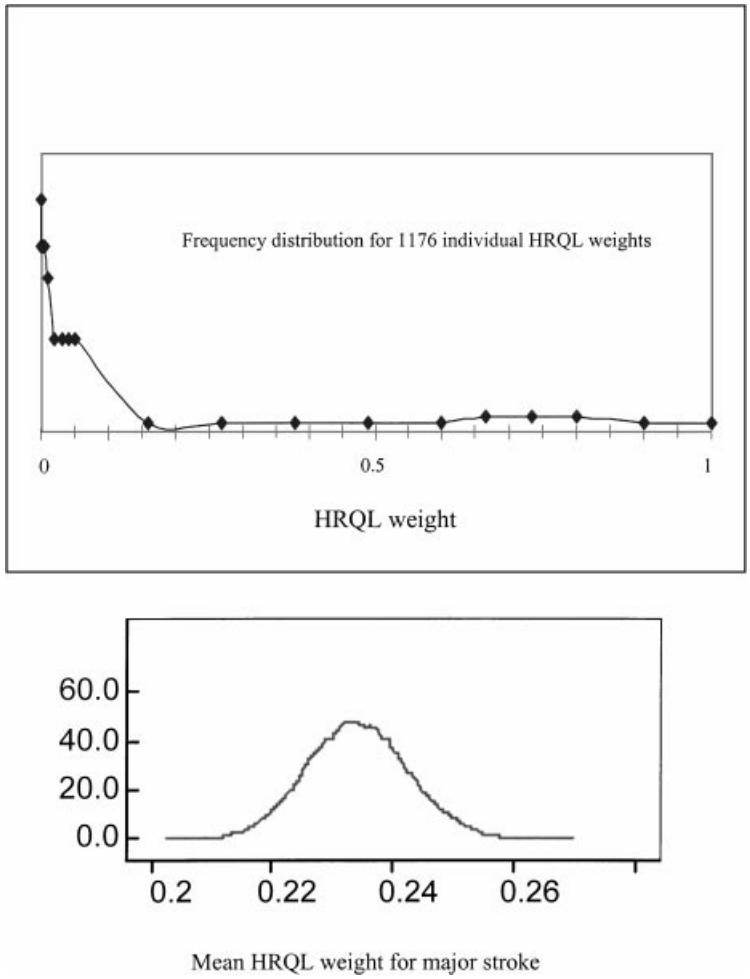

Figure 1. Upper panel shows the distribution of 1,176 estimates of HRQL weight elicited from persons with major stroke or at high risk of major stroke (personal communication, Dr. Greg Samsa, 1998). Lower panel shows estimated posterior density function for the mean HRQL weight where the prior on HRQL was the noninformative distribution beta(10^{-6},10^{-6}). This was updated based on data in the upper panel using Markov Chain Monte Carlo simulation (19,500 samples) in WinBUGS.

group; e) additional costs of stroke in the first year for patients who suffer disabling stroke within the 30 days of treatment; and f) annual costs of stroke in subsequent year for patients with stroke. Means for the first four of these costs are reported by Mark et al. (16) in the GUSTO CEA; however, no information about variance is provided. Costs for major stroke month by month for 36 months following an index hospitalization for stroke are reported by Lipscomb et al., who analyzed Medicare costs for stroke (13). Subsequent to that report, the Medicare cost data were supplemented by the same researchers with additional data about attributable costs not covered by Medicare (data provided by personal communication, Drs. G. Samsa and J. Lipscomb, 1998). In the GUSTO CEA, the hospitalization costs for stroke were included in the first-year baseline costs with each treatment group. However, the additional costs, such as those collected above the Medicare costs by the stroke PORT researchers, were not included in the published CEA. The costs of stroke beyond year 1 were not in the published base case CEA, but were included in one sensitivity analysis. We

https://doi.org/10.1017/S0266462301104083 Published online by Cambridge University Press

Table 3. Parameters for Normal Distributions for Mean Costs Used in Our Analysis and Sources of the Estimates

Item	Mean	SD (Precision[a])	Source
Cost for t-PA (per patient)	$2,750	95.3 (0.00011)	Mean from reference 16; precision, our estimate
Cost for SK (per patient)	$320	15.1 (0.0044)	Mean from reference 16; precision, our estimate
First-year medical costs, averaged across all t-PA patients (regardless whether survived past 30 days)	$24,990	200 (0.000025)	Mean from reference 16; precision, our estimate
First-year medical costs, averaged across all SK patients (regardless whether survived past 30 days)	$24,575	200 (0.000025)	Mean from reference 16; precision, our estimate
Costs of disabling stroke not captured in hospital costs in first year	$16,855	151 (0.000044)	Data from Samsa G, Lipscomb J, personal communication, 1998.
Annual costs of disabling stroke (after year 1)	$20,158	186 (0.000029)	Data from reference 13, with additional data provided by Samsa G, Lipscomb J, personal communication, 1998.

[a] WinBUGS parameterizes the normal distribution with precision, which is the reciprocal of the variance.

included both of these costs in our analysis. The assumed parameters for cost distributions used in our computations are given in Table 3.

If we had access to primary cost data from individual patients, how could we proceed? Costs (treated as a positive number) are bounded below by zero, and generally skewed with a long tail to the right. A useful distribution for describing these sorts of data is the gamma distribution. (The lognormal distribution has similar properties and has also been used to analyze costs in the same fashion as we use here with the gamma.) The gamma distribution is a two-parameter distribution, gamma (a, b) with mean a/b and variance a/b^2. The parameters are real numbers that must be greater than zero. A hierarchical Bayesian analysis can use cost data to update distributions on each of the two parameters for the gamma(a, b), which in turn induces a posterior distribution on the mean, a/b. It is this posterior distribution that we use in our CEA calculations. For this pedagogic analysis, we demonstrate the hierarchical approach and assume a relatively noninformative prior distribution for each of the parameters, a and b; thus, each of the parameters for the gamma distribution to describe cost is treated in turn as a random variable with its own prior distribution. (It is from this hierarchy of distributions on the parameters of distributions—the top level prior is called a hyper prior—that hierarchical analysis gets its name.) For these hyper prior distributions, we regularly use a gamma distribution with the WinBUGS default parameters of $a = b = 0.001$, which produces a very flat gamma density, not giving much likelihood to any particular value of the random variable. The mean of this distribution is 1, and the variance is 1,000; since a gamma distribution is bounded below at 0, this indicates an extremely long tail to the right, describing very well our prior uncertainty about each of the parameters of the gamma density for cost—they more or less can be anything greater than zero. A noninformative prior in turn means that the inferred parameters for the gamma distribution on costs are determined mostly from the data. We then use WinBUGS and Markov Chain Monte Carlo

https://doi.org/10.1017/S0266462301104083 Published online by Cambridge University Press

simulation to obtain estimated posterior distributions for each of the parameters, *a* and *b*, of the gamma distribution for individual costs.

However, our interest in cost-effectiveness analysis is the resulting posterior distribution on the mean cost, a/b (10). The posterior on this mean cost is inferred by the induced posterior on the ratio of the two random variables that parameterize the cost distribution. In cost-effectiveness analysis the ICER is a ratio of differences in means—differences in mean effectiveness and differences in mean costs, not a mean of observed costs and observed effectiveness measures—so it is the distribution on the mean costs in which we are interested (21). This distribution is not necessarily normal in the same sense that the sampling distribution of a sample mean is normal as a consequence of the central limit theorem of statistics. The mean in which we are interested is not the sum of sampled variables, but a true value for a parameter. Although there is no closed form expression for the posterior on the mean cost, even for moderate sample sizes for observed cost data it appears approximately normal. (Keep in mind that the distribution of costs is very skewed, but the distribution representing uncertainty about *mean* cost may not be very skewed.)

Since mean costs are reported in the GUSTO CEA without indication of their dispersion, we have assumed the posterior distributions of means of costs can be approximated by normal distributions with the reported means and variances representative of typical cost data for hospitalizations to which we have access and have experimented with Bayesian analysis. Although the normal distributions in Table 3 might at first thought be considered sampling distributions for sample means, the chain of reasoning leading to them was quite different.

Inferences About Long-term Survival. Because the GUSTO trial observed only 1-year survival, longer term survival had to be estimated by Mark et al. (16) in their CEA. They used the Duke Cardiovascular Disease Database to fit a survival function for 14-year survival, given the patient survives the first year, and then a Gompertz function to fit the tail survival curve reaching zero at approximately 35 years after the MI (16,1419). It is quite possible, given access to individual survival data, to use Bayesian analysis to compute posterior distributions for the parameters of hazard functions and thereby obtain induced posterior distributions on survival and life expectancy (19). However, we did not have access to primary data, so our analysis used only the point estimates of life expectancy provided in the published article (16,1420, interpolated from Figure 1). As did Mark et al., we discounted the survival function at a rate of 5%. In this fashion, we computed the present value of life-years survived past the first year to be 10.36, regardless of the arm of the trial (i.e., the only survival difference between patients treated with t-PA and with SK occurs in the first year).

A reviewer of this manuscript has noted we could have directly parameterized life expectancy for our calculations, put a prior on this parameter, and carried it through the final computations. This is indeed another way to deal with uncertainty about life expectancy. It would be necessary to discount life expectancy for the CEA computation, but the posterior on life expectancy would induce a posterior on discounted life expectancy. Because of limited time to prepare this manuscript, we have elected not to propagate uncertainty about life expectancy past the initial year in this model. The variance of our reported posterior distribution on net benefits will underestimate the variance of this distribution had we incorporated this added source of uncertainty.

Computational Methods

To this point we have specified distributions on the θ_i input parameters. Each distribution represents our uncertainty about a specific parameter, given what data we have access to, or our best judgment in the absence of data. Because the data come from many sources,

we cannot estimate the true *joint* posterior distribution on the θ_i (the input parameters for the analysis discussed earlier). We approximate the joint distribution under the assumption that the θ_i are independent random variables, and convolute the uncertainty from each into a single posterior distribution for each of the CEA computational results.

Ideally the independence assumption would not be needed. If data were available on as many of the θ_i as possible tied together in one observational unit at the individual patient level, then any correlational structure among the θ_i could be incorporated in the Bayesian estimates of the joint distribution for the final results. In this fashion, for example, a possible relationship between quality of life, survival time, and initial year hospital costs could be explicitly modeled and accounted for in the final joint posterior distribution on incremental costs and incremental QALYs. Demonstration of this type of modeling will await availability of data of this sort.

To combine the posterior uncertainty about each of the four proportions, six costs, and two quality-of-life weights into the calculations, we used Monte Carlo simulation with WinBUGS version 1.2 to calculate results for our recapitulation of the GUSTO CEA. This use of WinBUGS—as compared to its use to derive posterior distributions for some of the individual parameters—is not uniquely Bayesian. Rather, it is being used for probabilistic sensitivity analysis simulation (5). These WinBUGS results are reported here.

Figure 2 shows a WinBUGS doodle for convoluting two random variables into a function of their values. The node labeled "SKQalys" is a random variable that is a logical function of four predecessor random variables and a constant. (See Fryback DG, Stout NK, Rosenberg MA, "An Elementary Introduction to Bayesian Computing Using WinBUGS", in this issue, for terminology.) The predecessor variables are:

- SKsurv, the proportion of SK-treated patients who survive the first 30 days; this is represented as a beta density, as described earlier, with parameters A2 and B2 shown as rectangles. Numerical values for A2 and B2 were input in a data statement;

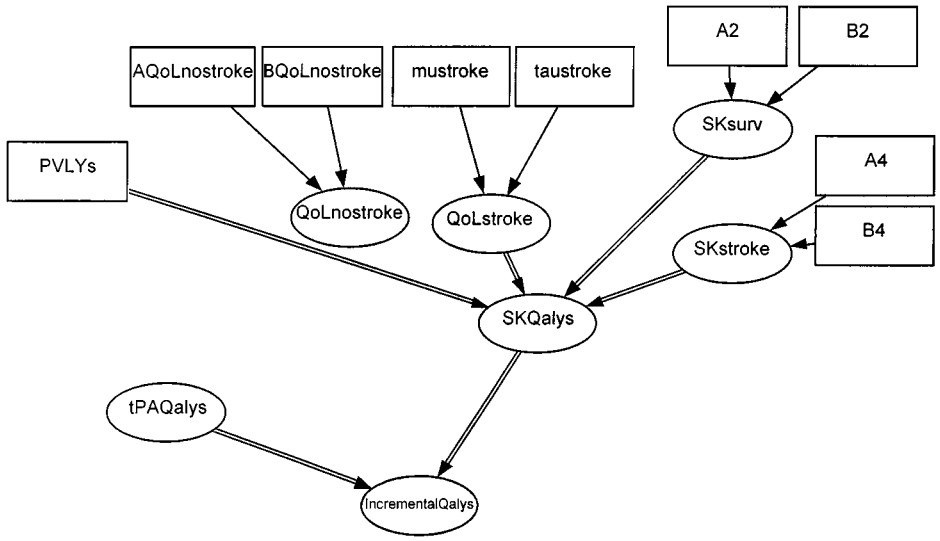

Figure 2. This partial WinBUGS doodle represents the computation to compute a distribution for QALYs experienced on average by patients treated with SK. This simulation convolutes distributions for the input variables into a distribution on the computed value of the variable, SKQalys. A similar portion of the full doodle showing the antecedents of tPAQalys has been suppressed in this figure for clarity.

https://doi.org/10.1017/S0266462301104083 Published online by Cambridge University Press

- SKstroke, the proportion of 30-day survivors among SK patients who have a nonfatal, disabling stroke in the first 30 days; another beta density, described earlier, with parameters input as constants A4 and B4;
- QoLstroke, the mean quality-of-life weight for a year with disabling stroke; this is represented by a normal density function, with constants mustroke and taustroke, approximating the empirical posterior density described earlier;
- QoLnostroke, the mean quality-of-life weight for patients without stroke, represented by a beta density with parameters AQoLns and BQoLns, as discussed earlier; and
- PVLY is a constant, calculated as the present value of the survivor curve past the first year, discounted continuously at 5%.

The computation of SKQalys is:

$$\text{SKQalys} = \text{PVLY}(\text{SKsurv}^{*}(\text{SKstroke}^{*}\text{QoLstrok} + (1 - \text{SKstroke})^{*}\text{QoLnostroke})).$$

WinBUGS accomplishes this by simulated random sampling (known as Monte Carlo simulation) from the distribution of each variable on the right side of the equation, conducting the computation with the sampled values, then accumulating a set of computed values for SKQalys across repeated sampling in this fashion. We used 20,000 samples. Figure 3 shows the resulting posterior distribution for SKQalys. A similar distribution was obtained for QALYs accumulated by t-PA treated patients. Finally, SKQalys was subtracted from tPAQalys to obtain a distribution on the incremental mean QALYs for t-PA compared with SK treatment.

With a similar process we used WinBUGS to simulate a distribution for mean incremental costs of t-PA treatment over SK treatment, using the distributions for costs discussed earlier. Our point estimate of the incremental cost-effectiveness ratio is the mean of the distribution for incremental costs, $3048, divided by the mean of the distribution for incremental QALYs, 0.13 QALY, which gives $23,446/QALY. The most comparable figure from the GUSTO CEA analysis is $32,678/QALY (16,1422, Table 3). Our calculation includes added costs of stroke and a quality of life for patients with disabling stroke not included in the published analysis.

Because the distribution of mean incremental QALYs considerably overlaps zero, a distribution for the incremental cost-effectiveness ratio, which divides mean incremental cost by mean incremental QALYs, is not well behaved where the denominator goes to zero.

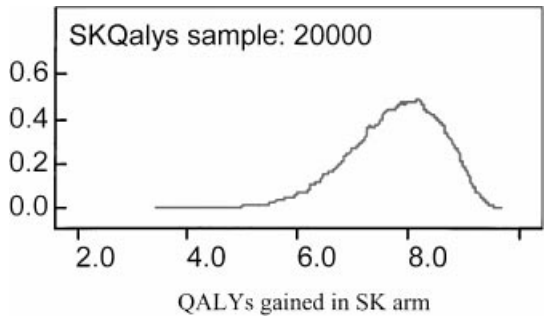

Figure 3. The distribution derived from 20,000 Monte Carlo samples, as described in the text, for QALYs estimated to accrue on average to patients treated with SK. The mean is 7.75 QALYs, and the central 95% credible interval ranges from 5.86 to 9.11 QALYs. Small irregularities in the smoothness of the density are due to the sampling process and are inconsequential.

https://doi.org/10.1017/S0266462301104083 Published online by Cambridge University Press

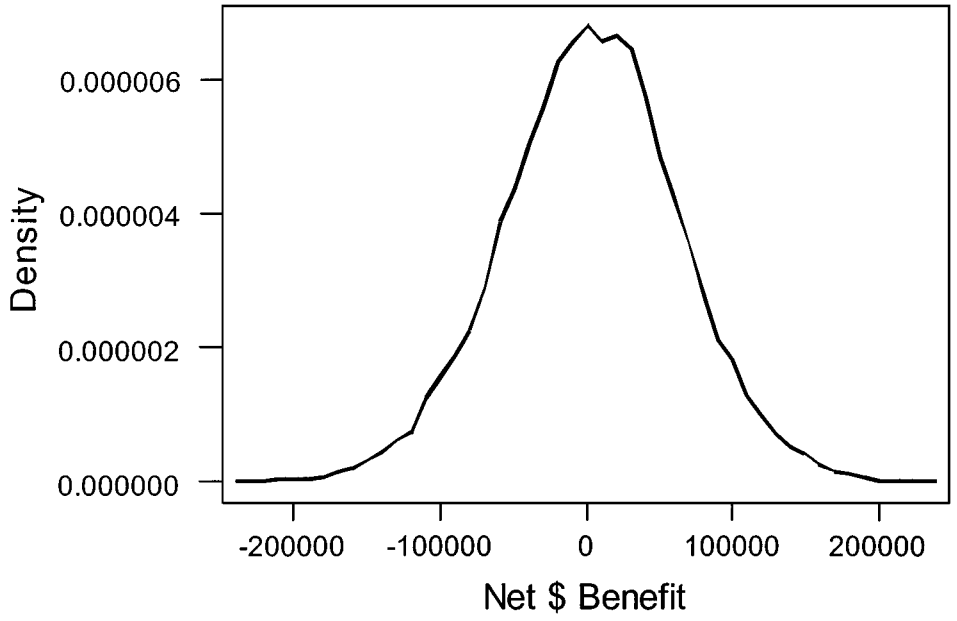

Figure 4. The distribution for net benefit of t-PA compared with streptokinase at a trade-off rate of $50,000 per QALY, using the distributions for input parameters described in the text. The best point estimate for the net benefit is the mean of this distribution, $3,339; the central 95% credible interval ranges from −$116,490 to $123,040. A value of $0 for this quantity means that the incremental cost-effectiveness ratio is $50,000/QALY, and values above $0 (positive benefit) indicate a cost/QALY ratio below (i.e., more favorable than) $50,000. The probability the incremental net dollar benefit is above $0 is 52.5% according to this simulated posterior density. Small irregularities arise from the sampling process and are inconsequential.

Similarly, if the incremental QALY is positive and the incremental mean cost is negative, computation of the ratio is meaningless since one treatment dominates the other. For these reasons, we have avoided dealing with a posterior distribution on the ICER. Instead, we elected to compute a posterior distribution for net benefit, converting QALYs to dollars at a rate of $50,000 per QALY, using methods proposed by Stinnett and Mullahy (20). The estimated distribution for net benefit at this trade-off rate is shown in Figure 4. The mean of this distribution is $3,339, and 52.5% of it is above a net benefit value of $0. If we compute a comparable figure from the study by Mark et al. (16) for their base case analysis, where the average QALY gain in present value was (0.09 life years) × 0.9 (QALY per LY) = 0.081QALY at an average incremental cost of $2,845 with the present value of $2,647, the average net benefit was (0.081 QALY) × (50,000 $/QALY) − $2,647 = $1,403.

Our figures differ. We believe there are two sources for this difference. First, incorporating randomized trial data in the prior for SK but not for t-PA depressed the overall estimate of SK effectiveness compared with t-PA. The prior trials found no difference between SK and t-PA 30-day survival, both of which were estimated to be about 90%. The GUSTO trial resulted in data with 93.7% for t-PA and 92.7% for SK. In Table 2 it is seen that the differential manner in which we incorporated prior data for estimating the two proportions accentuated the effectiveness of t-PA by depressing effectiveness of SK by partially incorporating the lower survival rate in earlier trials. If we re-run the analysis using beta (0,0) as the prior for all four estimates (i.e., giving no weight to prior trial data at all and using the GUSTO data only), the estimated mean net benefit is $258 with a somewhat larger fraction, 50.2%, of the posterior distribution below zero. This is roughly congruent with the result

https://doi.org/10.1017/S0266462301104083 Published online by Cambridge University Press

by Mark et al. (16) but still implies an ICER higher than their result. The residual difference is due to the fact that the mean quality of life in our analysis is somewhat lower than the 0.9 assessed from the GUSTO patients alone, and the fact that we incorporate more costs of stroke, albeit with low probability.

Of most interest is the uncertainty in the net dollar benefit estimate. Although the mean is shifted higher using prior information less favorable to SK than in the GUSTO study, the probability that the mean net benefit is above zero—i.e., that the ICER is less than $50,000 per QALY—is only slightly higher than 50%. This is much more uncertain than the traditional CEA by Mark et al. (16) connoted. This could be pursued using the same inputs as we used. However, we believe time is better spent pursuing a full Bayesian analysis of GUSTO from primary data than spent analyzing the final residual differences between our pedagogic analysis and the results by Mark et al. (16).

DISCUSSION

Bayesian analyses are beginning to appear in the healthcare cost-effectiveness literature (2;12), although the genre is still rare. The computational challenge is considerable but is made easier with Markov Chain Monte Carlo software such as WinBUGS.

There are a number of intellectual challenges to be solved to truly make the approach informative. First, viable models for combining data from many sources into Bayesian posterior distributions need to be developed. A good start on this was achieved by Eddy and colleagues a decade ago (7;8). They developed likelihood models for a variety of experimental and observational study designs, incorporating corrections for various potential biases. Analyses of this sort will need to be pursued to estimate the parameters needed for CEA models based on secondary analyses.

Second, analysts using Bayesian techniques need to develop CEA models from data on costs, effect, and quality of life collected at the individual patient level in order to incorporate the correlational structure that surely exists among them. Developing models dealing with this structure is a definite challenge.

Third, Bayesian analysts will need to explore when and how to model prior information. Incorporation of prior information is a source of both strength and criticism for the Bayesian approach. It is a strength because prior data is obviously relevant and in fact should be used to inform inferences about costs and effectiveness—witness the growing popularity of meta-analyses. It is a source of criticism because the ultimately subjective decisions about what goes into a prior and how much it is weighted are seen as a means of manipulating results. The research paradigms for resolving the apparent conflict in these points of view have yet to be developed in this application.

We believe that Bayesian analysis offers a useful approach to dealing with the uncertainties in CEA modeling. We agree with Claxton (4) that conquering the computational challenges of Bayesian analyses in CEAs allows us to go to the next step of considering the value of additional information in policy decisions using CEA results.

REFERENCES

1. Briggs A, Sculpher MJ, Buxton MJ. Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. *Health Econ*. 1994;3:95-104.
2. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ*. 1999; 8:257-261.
3. Brophy JM, Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *JAMA*. 1995;273:871-875.
4. Claxton K. The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ*. 1999;18:341-364.

5. Doubilet P, Begg CB, Weinstein MC, et al. Probabilistic sensitivity analysis using Monte Carlo simulation: A practical approach. *Med Decis Making*. 1985;5:157-177.
6. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press; 1997.
7. Eddy DM, Hasselblad V, Shachter R. A Bayesian method for synthesizing evidence. The confidence profile method. *Int J Technol Assess Health Care*. 1990;6:31-55.
8. Eddy DM, Hasselblad V, Shachter R. An introduction to a Bayesian method for meta-analysis: The confidence profile method. *Med Decis Making*. 1990;10:15-23.
9. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: Initial catalog of health-state quality factors. *Med Decis Making*. 1993;13:89-102.
10. Gilks WR, Richardson S, Spiegelhalter DG. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall; 1996.
11. The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med*. 1993;329:673-682.
12. Johnson-Masotti AP, Laud PW, Hoffmann R, et al. Bayesian analysis applied to cost-effectiveness data [abstract]. *Med Decis Making*. 1999;19:542.
13. Lipscomb J, Ancukiewicz M, Parmigiani G, et al. Predicting the cost of illness: A comparison of alternative models applied to stroke. *Med Decis Making*. 1998;18(suppl):S39-S56.
14. Mandelblatt J, Fryback D, Weinstein M, et al. Assessing the effectiveness of health interventions for cost-effectiveness analysis. Panel on Cost-Effectiveness in Health and Medicine. *J Gen Intern Med*. 1997;12:551-558.
15. Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR, Siegel SJ, Russell LB, Weinstein MC, eds. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996.
16. Mark DB, Hlatky MA, Califf RM, et al. Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. *N Engl J Med*. 1995;332:1418-1424.
17. Neuhaus KL, von Essen R, Tebbe U, et al. Improved thrombolysis in acute myocardial infarction with front-loaded administration of alteplase: Results of the rt-PA-APSAC patency study (TAPS). *J Am Coll Cardiol*. 1992;19:885-891.
18. O'Brien BJ, Drummond MF, Labelle RJ, et al. In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care*. 1994;32: 150-163.
19. Stangl DK. Hierarchical analysis of continuous-time survival models. In: Berry DA, Stangl DK, eds. *Bayesian biostatistics*. New York: Marcel Dekker; 1996:429-450.
20. Stinnett AA, Mullahy J. Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making*. 1998;18:S68-S80.
21. Stinnett AA, Paltiel AD. Estimating CE ratios under second-order uncertainty: The mean ratio versus the ratio of means. *Med Decis Making*. 1997;17:483-489.
22. Tsevat J, Goldman L, Soukup JR, et al. Stability of time-tradeoff utilities in survivors of myocardial infarction. *Med Decis Making*. 1993;13:161-165.