

Bayeswatch: an overview of Bayesian statistics

Peter C. Austin PhD,^{1,2} Lawrence J. Brunner PhD PhD³ and Janet E. Hux MD SM FRCP(C)^{1,4,5}

¹Institute for Clinical Evaluative Sciences, Toronto, Canada

²Department of Public Health Sciences, University of Toronto, Toronto, Canada

³Department of Statistics, University of Toronto, Toronto, Canada

⁴Department of Medicine, University of Toronto, Toronto, Canada

⁵Division of General Internal Medicine, Clinical Epidemiology Unit and Health Care Research Program, Sunnybrook and Women's College Health Sciences Centre, Toronto, Canada

Correspondence

Dr Peter Austin
Institute for Clinical Evaluative Sciences
G-160, 2075 Bayview Avenue
North York, Ontario
M4N 3M5
Canada
E-mail: peter.austin@ices.on.ca

Keywords: Bayesian statistics,
statistical methodology, statistics

Accepted for publication:
24 November 2001

Abstract

Increasingly, clinical research is evaluated on the quality of its statistical analysis. Traditionally, statistical analyses in clinical research have been carried out from a 'frequentist' perspective. The presence of an alternative paradigm – the Bayesian paradigm – has been relatively unknown in clinical research until recently. There is currently a growing interest in the use of Bayesian statistics in health care research. This is due both to a growing realization of the limitations of frequentist methods and to the ability of Bayesian methods explicitly to incorporate prior expert knowledge and belief into the analyses. This is in contrast to frequentist methods, where prior experience and beliefs tend to be incorporated into the analyses in an *ad hoc* fashion. This paper outlines the frequentist and Bayesian paradigms. Acute myocardial infarction mortality data are then analysed from both a Bayesian and a frequentist perspective. In some analyses, the two methods are seen to produce comparable results; in others, they produce different results. It is noted that in this example, there are clinically relevant questions that are more easily addressed from a Bayesian perspective. Finally, areas in clinical research where Bayesian ideas are increasingly common are highlighted.

Introduction

The evaluation of clinical research findings increasingly rests on the statistical rigour of the analysis, rather than on the clinical importance of the results. Traditionally, statistical analyses in clinical research have been carried out from a 'frequentist' perspective, in which a hypothesis of no effect is rejected if a *P*-value is sufficiently small. The existence of a competing paradigm, the Bayesian paradigm, has been relatively unknown in clinical and health care

research. However, recent interest in, and application of, Bayesian methods is growing rapidly.

Bayesian methods allow one explicitly to incorporate existing knowledge and expert opinion into the analysis of data. They also allow one to ask different questions of the data than are possible from the frequentist perspective. Goodman (1999a, 1999b) argues for the superiority of Bayesian methods for analysing clinical data generally. Lilford & Braunholtz (1996) argue against the use of conventional statistical tests as a basis for implementing policy

based on clinical research, stating a concern that conventional statistical tests dichotomize results according to whether they are significant or not, forcing what they describe as an 'off/on' response from decision-makers.

One reason Bayesian methods may be preferred over the frequentist approach is that they more closely approximate our natural thought processes (Freedman 1996; Davidoff 1999; Malakof 1999). The frequentist approach to statistics is deductive (Freedman 1996; Davidoff 1999; Goodman 1999a): it assumes that a given hypothesis is true, and then calculates the probability of observing an outcome at least as extreme as that which was observed. The resultant probability is called the *P*-value. The Bayesian paradigm is an inductive approach to inference, which, given observed data, allows one to assess the likelihood of a given hypothesis.

This paper begins with an introduction to both the frequentist and Bayesian approaches to statistics, followed by an example in which acute myocardial infarction (AMI) mortality data are analysed from both a frequentist and a Bayesian perspective. The results of the two analyses are compared and the interpretations contrasted. Finally, some applications of Bayesian methods in clinical research are outlined.

Background to frequentist and Bayesian methods

Frequentists and Bayesians use the term 'probability' in different ways. The traditional approach to statistical inference is called 'frequentist' because of the way it interprets probability. When a random event (like tossing a coin) is repeated a large number of times independently and under identical conditions, the probability of an event is approached by way of its relative frequency of occurrence. Hence, for the frequentist, probability is a model of long-run relative frequency. Note that with this interpretation, it is illogical to speak of the probability that it will rain tomorrow, or that a particular patient will be alive 5 years from now. One would have to imagine a series of tomorrows, all occurring under identical circumstances, and observe in what proportion of those tomorrows rain occurred.

The Bayesian approach takes its name from the English clergyman Thomas Bayes (1702–1761), who

was responsible for a theorem specifying certain relations among conditional probabilities. Bayes' theorem says that the probability distribution of a given parameter, conditional on the observed data, is equal to the product of the prior probability distribution of the parameter with the likelihood function, divided by the probability of the data. In the Bayesian approach to statistical inference, probability is a model of scientific knowledge. As such, there is a subjective element to probability, in that different observers may have differing degrees of belief in the likelihood of a specific event or hypothesis. Still, it is assumed that these beliefs are constrained by the usual rules of probability. However, there is evidence (French & Smith 1997) that informal human judgements do not strictly conform to the laws of probability. Expert opinion or belief that exists prior to observing the data can be converted into a probability distribution called a prior probability distribution. A model for observable data, given that some 'fact' is true, is combined with the prior probability distribution via Bayes' theorem to yield a posterior probability that the 'fact' is true, given the observed data. This posterior probability represents the state of knowledge after observing the data, and may be used as prior information in subsequent studies. Therefore, one can say that the Bayesian approach to statistical analysis is an idealized model of how subjective belief ought to change when a rational being encounters new data. Some argue that, for a Bayesian, scientific knowledge emerges not from objective analyses but from consensus: when a majority of scientists, each starting from their respective prior beliefs, are led to the same inference on the basis of the data (French & Smith 1997). One of the great strengths of the Bayesian method is that it allows expert knowledge, in the form of a prior probability distribution, to be formally incorporated into the statistical analysis. In contrast to this, the use of prior information in frequentist inference tends to be informal and unsystematic.

The differences between Bayesian and frequentist methods are illustrated in their treatment of constants and random variables. A random variable is a quantity that can assume any of a number of possible values, each with a given probability or likelihood. For instance, in rolling a fair die, the number that will come up is a random variable,

taking the numbers 1 to 6, each with probability 1 in 6.

A constant can only assume one value. From a frequentist perspective, the name of the first author of this paper is a constant, since it can only take one value. There is not a set of values that the name could potentially assume. One can make probabilistic statements only about random variables. Given a fair die, one can ask what is the probability that upon rolling it, the number one turns up. In contrast, one cannot make probabilistic statements about constants. From a frequentist perspective, one cannot ask what is the probability that the first author of this paper is called 'Dave', since he is not called 'Dave'.

The second author will now comment on the previous paragraph. The statements the first author makes about his name are correct, from his subjective perspective. However, whether the first author's name is 'Dave' is in fact a random variable. In the mind of the second author, there is some infinitesimally small probability that the first author really is named 'Dave'. Perhaps he is leading a double life, or is a spy under very deep cover, or is being shielded as part of a witness protection programme. In the mind of someone that has never heard of the first author and has not yet seen this paper, the probability that the first author is named 'Dave' may be some appreciable fraction of one percent. Now, upon seeing the first page of this paper, such a person's subjective probability of 'Dave' should dip to near zero, while the second author's beliefs will not be affected at all.

The frequentist approach to statistics (Cox & Hinkley 1974) assumes that parameters (e.g. means, variances and regression coefficients) are unknown constants characterizing the larger population from which the data were drawn. The analyst seeks to generate estimates of these true (but unknown) population parameters, and computes sample statistics accordingly. Given observed data, and a model containing parameters, the likelihood function (Fraser 1976) is the likelihood of observing the given data, conditional on a particular set of parameter values. Statistical estimation is often carried out using maximum likelihood methods, in which the parameter estimates are those parameter values under which the observed data were most likely to have arisen. Given observed data, one can formulate a null

hypothesis concerning the underlying model through which the data were generated. The null hypothesis is typically one of no effect. Associated with each null hypothesis is a *P*-value (Cox & Hinkley 1974). This is the probability of obtaining an outcome at least as extreme as that observed, under the assumption that the null hypothesis is true. The *P*-value is a rough measure of the consistency of the data with the null hypothesis. Royall (1997) criticized the use of the *P*-value for two reasons. Firstly, it is confounded with sample size. Secondly, it refers to unobserved data ('results at least as extreme as that observed'). The reference to that which is unobserved violates the likelihood principle (Royall 1997) – that conclusions should only be based on the observed data.

Neyman and Pearson (1933) developed a formal hypothesis-testing framework. Any statistical test of hypothesis has two components: the null hypothesis (e.g. the mean is zero) and an alternative hypothesis (e.g. the mean is not zero). For each hypothesis test, a test statistic is determined. If the test statistic exceeds a predetermined threshold, the null hypothesis is rejected. The threshold is chosen so as to limit the rate of rejecting the null hypothesis to a pre-specified level when the null hypothesis is indeed true (type I error), and to limit the rate of accepting the null hypothesis when the alternate hypothesis is true (type II error). The outcome of a hypothesis test is to be a decision: to accept or reject the null hypothesis. Contrary to common belief, one does not either reject or fail to reject the null hypothesis – one either accepts or rejects a given hypothesis (Neyman 1950). The threshold at which one accepts the alternate hypothesis is chosen to limit the long-term rate of incorrect decisions. Neyman (1950) comments that the result of a hypothesis test is to be an action. Goodman (1999a) notes that the result of a hypothesis test is to be a decision, and not an inference, and that there is no number that reflects back to the strength or weakness of the evidence for or against the specific hypothesis of interest.

Frequentist methods permit the construction of confidence intervals around parameter estimates. Whilst it is tempting to simplify the explanation of a 95% confidence interval to an interval such that there is a 95% chance that the true parameter value lies in the interval, the frequentist must reject such a formulation since it implies that the true parameter

value is a random variable and not a fixed quantity. From the frequentist perspective, only the data are random, whereas the parameters are fixed but unknown quantities. For a frequentist, then, a 95% confidence interval is an interval such that if data were generated repeatedly under identical conditions, then 95% of the constructed confidence intervals would contain the true (but unknown) parameter value. The 95% confidence interval only tells us about the precision with which we have estimated the underlying parameter, and nothing of its location. Similarly, one cannot talk about the probability that a parameter exceeds a specified threshold. The population parameter of interest is a fixed value, and either exceeds the threshold, or does not.

The Bayesian paradigm (Cox & Hinkley 1974; Lee 1989; Gelman *et al.* 1995) views both the data and the underlying parameters (means, regression coefficients, variances, etc.) that generated the data as random variables – random because they are unknown. The Bayesian perspective allows one to explicitly incorporate prior beliefs and expert knowledge concerning underlying parameters with the observed data to obtain probability distributions of the parameters. The prior probability distribution summarizes all available information and expert opinion concerning the parameter of interest before the data have been observed. Bayes' theorem provides a method of combining the likelihood function with the prior probability distribution to obtain the posterior probability distribution. The posterior probability distribution is the probability distribution of the unobserved parameter, conditional on the observed data, given one's prior beliefs. For a Bayesian, all knowledge about the underlying parameter, after observing the data, is contained in the posterior probability distribution.

Once the posterior distribution has been determined, one is able to make probabilistic statements about the underlying, unobserved parameters, such as the probability that a given parameter exceeds a specified threshold, or that the parameter of interest lies in a certain interval. Bayesian methods allow the construction of 'credible intervals' for parameters of interest. A 95% credible interval is an interval such that the true population parameter lies in the interval with probability 0.95. Therefore, a 95% credible

interval provides the most likely location of the underlying parameter.

The Bayesian model explicitly combines prior beliefs concerning the parameter's probability distribution with the likelihood function to produce a posterior distribution. As the sample size increases, the likelihood dominates the prior (Gelman *et al.* 1995). In such situations, the choice of prior probability distribution will have little influence on the posterior probability distribution. That is, two scientists who may start out with very different prior opinions may come to different conclusions based on a small amount of data. However, their conclusions will converge as the amount of data increases.

The Bayesian paradigm permits hypothesis testing, using Bayes factors. Suppose one wishes to compare the evidence for a null hypothesis vs. an alternative hypothesis. The Bayes factor is defined as a ratio: the probability of obtaining the observed data given the null hypothesis, divided by the probability of obtaining the observed data under the alternative hypothesis. Naturally, small values are taken as evidence against the null hypothesis, since they mean that the data are relatively more likely to have occurred, given the alternative hypothesis. If one of the hypotheses is a composite hypothesis, then the probability of the data is the weighted average of the likelihood of the data given each of the possible components of that hypothesis, with the weights derived from the prior probability. Bayes factors allow one to determine to what degree prior beliefs in the likelihood of the null hypothesis have been modified by the data. Guidelines exist for deciding how small the Bayes factor has to be before one chooses the alternative hypothesis over the null hypothesis (Goodman 1999b). Goodman suggests that a Bayes factor of 1/5, 1/10, 1/20 and 1/100 be taken as weak, moderate, moderate to strong and strong to very strong evidence against the null hypothesis, respectively. In general, the optimal choice of a cut-off depends upon the benefits of deciding correctly and the cost of deciding incorrectly.

Traditionally, it has been surprisingly difficult to compute posterior probabilities and Bayes factors, except in the simplest of cases. However, with the advent of Markov Chain Monte Carlo (MCMC) methods (Gilks *et al.* 1996), Bayesian methods are

being implemented with increasing frequency. MCMC methods are computer-intensive procedures that allow one to simulate draws from the posterior distribution without having to calculate the posterior distribution explicitly. The mean or specified quantiles of the posterior distribution can be estimated by computing the mean or specified quantiles of the simulated draws from the posterior distribution. The advent of MCMC methods is largely responsible for the increasing interest in Bayesian methods. Bayesian calculations that would have been largely intractable are now relatively simple to implement using MCMC methods.

Hypothetical case study

For our case study, we examine the mortality rate for one hospital's patients admitted for acute myocardial infarction (AMI). For the hospital in question, out of 100 patients in a given year, 10 patients died following admission, prior to hospital discharge. Assume that the overall observed in-hospital mortality rate in the population of AMI patients, at all hospitals and for all patients, is 15%. In the case study, we contrast a frequentist with a Bayesian analysis of the above data. The results of the case study are summarized in Table 1. We will ask the following questions of the data: firstly, is the given hospital different from average?; secondly, is the hospital better than average?; thirdly, what is the probability that the hospital delivers excellent care, when excellence in medical care is defined as a mortality rate of less than 10%?

In both the frequentist and Bayesian analyses, we assume that the number of deaths follows a binomial distribution. In the frequentist analyses, confidence intervals and *P*-values will be determined using large-sample normal approximations to the binomial distributions. In the Bayesian analyses, central or equal tail credible intervals will be constructed.

We begin with a frequentist analysis of the data. Figure 1 depicts the likelihood function for the hospital's true in-hospital mortality rate. The maximum-likelihood estimate of the hospital's mortality rate is 10%. This is the most likely estimate of the hospital's mortality rate for patients admitted for an AMI, given the observed data. The 95% confidence interval around the estimated mortality rate is (4.1%,

Table 1 Comparison of frequentist and Bayesian analyses of AMI mortality data

Question	Frequentist analysis	Bayesian analysis I (diffuse prior)	Bayesian analysis II (informative prior)
Most likely estimate of mortality 95% confidence (credible) interval	10.0% (4.1%, 15.9%)	10.0% (5.6%, 17.5%)	10.2% (5.9%, 17.0%)
H ₀ : mortality equals 15%	P = 0.16, accept null hypothesis	BF = 4.5, no evidence against null hypothesis	BF = 1.1. Very little evidence against null hypothesis
H ₀ : mortality is at least 15%	P = 0.08, accept null hypothesis	BF = 1/55, strong evidence against null hypothesis	BF = 1/9.1, moderate evidence against null hypothesis
Probability that mortality rate is below 10%	Not applicable	0.43, moderate posterior probability that mortality rate lies below 10%	0.41, moderate posterior probability that mortality rate lies below 10%

P denotes the *P*-value for a given hypothesis test. BF denotes the Bayes factor for comparing two hypotheses.

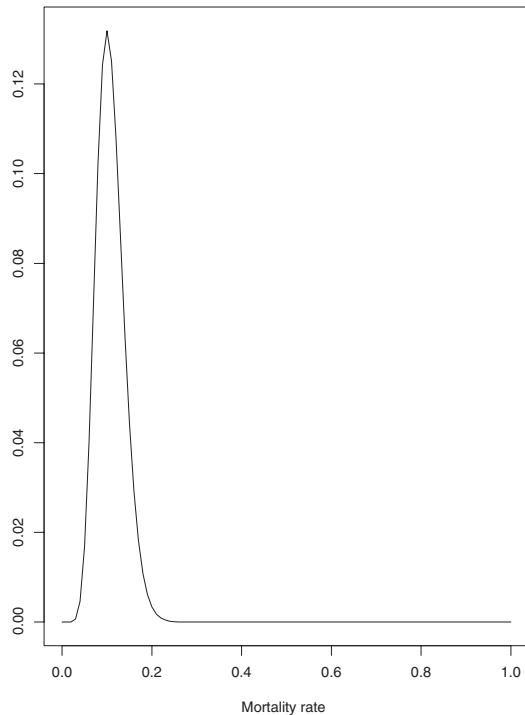


Figure 1 Likelihood function for AMI mortality data.

15.9%). If data were to be generated repeatedly, under identical circumstances for the same hospital, then 95% of the constructed confidence intervals would contain the hospital's true in-hospital mortality rate.

To address the first question ('is the hospital different from average?'), we formally test the null hypothesis that the hospital's mortality rate is equal to 15% against the alternate hypothesis that the mortality rate is different from 15%. The result would be to accept the null hypothesis that the hospital's true mortality rate is no different from the average mortality rate of 15% in the population of AMI patients ($P = 0.16$). To address the second question ('is the hospital better than average?'), one can test the null hypothesis that the hospital's true mortality rate is at least 15% against the alternative hypothesis that the true mortality rate is less than 15%. The P -value associated with this test is 0.08. Therefore, we would accept the null hypothesis that the hospital's true mortality rate is at least as high as the population average.

The frequentist paradigm does not allow us to address the third question. From the frequentist per-

spective, the hospital's mortality rate is a constant, and not a random variable. Therefore, it is either higher or lower than 10%, but one cannot assign a probability to either scenario.

The Bayesian paradigm allows us to avoid beginning with a specific hypothesis about the hospital's mortality rate. The Bayesian analysis begins by specifying a prior probability distribution for the hospital's in-hospital AMI mortality rate. This distribution may incorporate expert knowledge and specific information about the hospital being studied, such as the annual volume of AMI patients or academic affiliation of the hospital. The prior must be specified before the data are examined.

The first prior distribution that we will assume is a non-informative, or diffuse, prior distribution. This means that the hospital's mortality rate can take on any value from 0 to 1 with equal likelihood. By specifying this prior distribution, we are saying that our prior beliefs about the hospital's mortality rate are vague or diffuse. By assuming this prior distribution, we assume, for example, that the hospital's mortality rate is as likely to lie below 10% as it is to lie above 90% (an unlikely scenario).

Using Bayes' theorem, we combine the prior probability distribution and the likelihood function, which is derived solely from the data and the assumed probability model, to obtain the posterior probability distribution. This is the probability distribution of the hospital's mortality rate, conditional on the observed data, given our prior beliefs concerning the parameter's probability distribution. Figure 2 depicts the diffuse prior probability distribution and the derived posterior probability distribution. The mode of the posterior distribution gives the most likely value for the hospital's in-hospital mortality rate, given our prior beliefs, conditional on the observed data, as 10.0%. The most likely posterior value for the hospital's mortality rate is the same as the frequentist maximum likelihood estimate. This will always happen when a uniform prior distribution is assumed. The associated 95% credible interval is (5.6%, 17.5%). The hospital's true mortality rate lies in this interval with probability 0.95.

To address the first question, of whether the hospital is different from average, one can compute the Bayes factor comparing the null hypothesis that the hospital's true mortality rate is 15% vs. the alternate

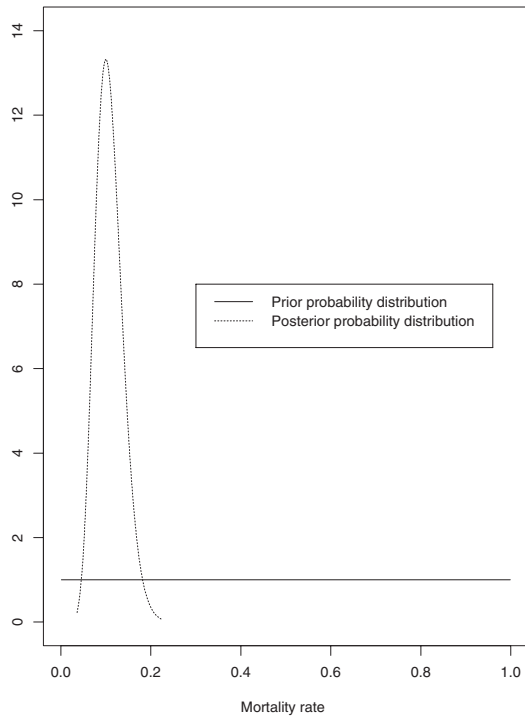


Figure 2 Non-informative prior beliefs.

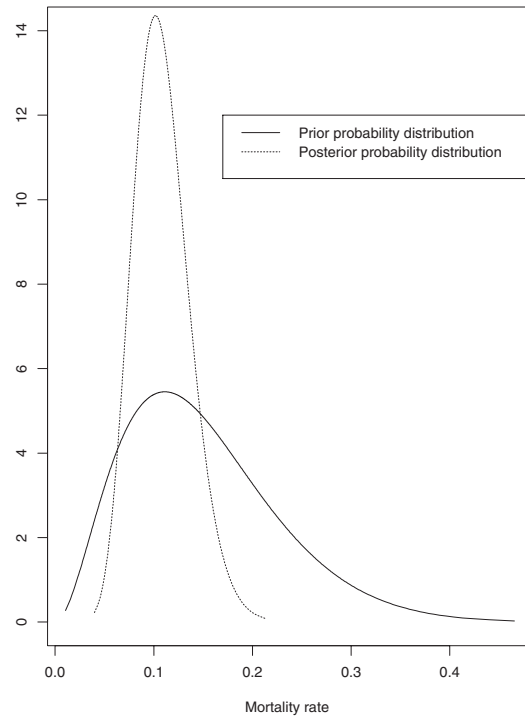


Figure 3 Informative prior beliefs.

hypothesis that the mortality rate is different from 15%. The Bayes factor associated with this test of hypothesis is 4.5. Since it is greater than one, there is no evidence in favour of the alternate hypothesis.

To address the second question, of whether the hospital is better than average, one can similarly compute the Bayes factor comparing the null hypothesis that the hospital's mortality rate is higher than average to the alternate hypothesis that the mortality rate is lower than average. The Bayes factor associated with this test of hypothesis is $1/55$. There is strong evidence in favour of the alternate hypothesis – that the hospital has a lower than average mortality rate.

Bayesian methods allow one to address the third question. In the Bayesian paradigm, the hospital's mortality rate follows a probability distribution. Hence, one can make probabilistic statements concerning the hospital's mortality rate. The probability that the mortality rate lies below 10% is 0.43. If a mortality rate of 10% denotes excellence in medical care, one is able to assess the probability that the hospital delivers excellent medical care. One can thus

use clinically informed criteria to assess the hospital's performance.

The above results depend on the prior probability distribution that was assumed at the beginning of the analysis. As a sensitivity analysis, one can modify assumptions concerning the prior probability distribution to see how the results vary. However, for large samples, the impact of the choice of prior probability distribution will disappear (Gelman *et al.* 1995). We repeat the analysis, assuming a more informative prior probability distribution. Let us assume that our prior beliefs about the hospital's mortality rate are worth the equivalent of observing the hospital treat 20 patients, and that we believe the mean in-hospital mortality rate to be 15% (the underlying mathematical form of the prior probability distribution has two parameters, requiring us to specify the above information). A prior distribution representing our beliefs and the posterior probability distribution are illustrated in Fig. 3. The mode of the hospital's mortality rate distribution is 10.2%. The associated 95% credible interval is (5.9%, 17.0%).

To address the first question, one can now test the

simple null hypothesis that the hospital's mortality rate is 15% vs. the alternate hypothesis that it is different from 15%. The Bayes factor associated with this test is 1/1.1. Therefore, there is very little evidence in favour of the alternate hypothesis over the null hypothesis. To address the second question, one can test the compound null hypothesis that the hospital's true mortality rate is at least 15% against the alternate hypothesis that it is less than 15%. The Bayes factor associated with this test is 1/9.1. Therefore, there is moderate evidence that the hospital's mortality rate is lower than average. The probability that the hospital's mortality rate is less than 10% is 0.41, allowing us to address the third question.

The sensitivity analysis showed that the hospital's estimated in-hospital mortality rate is relatively insensitive to choice of prior distribution. By varying our choice of prior distribution, the probability that the mortality rate lies below 10% is changed slightly. Under both prior probability distributions, there was no evidence in favour of the hypothesis that the hospital's mortality rate was different to the average. When evaluating the evidence that the hospital's mortality rate was lower than average compared with higher than average, the weight of evidence decreased with the more informative prior, from strong evidence to moderate evidence.

Discussion

In comparing the frequentist and Bayesian analyses of the case study data, one notices both similarities and differences in the conclusions that were drawn. The most likely value for the hospital's mortality rate was the same in the frequentist analysis as in the first Bayesian analysis. The most likely value obtained from the second Bayesian analysis was only marginally higher than that in the two other analyses. The 95% confidence interval and the 95% Bayesian credible intervals were of approximately the same length, although the two Bayesian intervals were shifted upwards compared with the frequentist confidence interval.

The frequentist analysis did not reject the null hypothesis that the hospital's mortality rate was different from average ($P=0.16$). Similarly, in both Bayesian analyses, there was little evidence that the

hospital was different than average. However, when we tested the null hypothesis that the hospital was worse than average against the alternate hypothesis that the hospital was better than average, the two paradigms reached strikingly different conclusions. The frequentist analysis did not reject the null hypothesis ($P=0.08$), whereas the two Bayesian analyses found moderate to strong evidence in favour of the alternate hypothesis. Strictly speaking, a frequentist should act as if the hospital was worse than average, whereas a Bayesian would have strong reason to believe that the hospital was better than average.

Finally, the two Bayesian analyses were able to assess the probability that the hospital's mortality rate was below specified thresholds. According to the Bayesian analysis, there was a moderate probability that the hospital had a mortality rate that was below 10%. From a frequentist perspective, this question cannot be addressed, since the hospital's mortality rate is seen as a constant, and therefore does not follow a probability distribution.

In the frequentist paradigm, the hospital's in-hospital mortality rate is seen as a fixed (but unknown) constant. The observed data provide the most likely estimate of the fixed (but unknown) mortality rate. One is able to test the hypothesis that the true mortality rate is equal to a specified value. Since the true mortality rate is a constant, and not a random variable, one cannot determine with what probability it exceeds a given threshold or lies in a specified interval. Frequentists interpret probability as the long-term rate of occurrence of an event. Hence, confidence intervals and P -values associated with hypothesis tests are interpreted as what would happen if data were generated repeatedly under similar circumstances.

In the Bayesian paradigm, the hospital's mortality rate is a random variable that follows a probability distribution. Data about the hospital's mortality rate are combined with prior beliefs concerning this probability distribution. Once the posterior probability distribution has been determined, one can make probabilistic statements concerning the hospital's mortality rate. If a mortality rate of 10% denotes excellence in medical care, one can determine the probability that the hospital's mortality rate lies below this threshold. In doing so, one can assess the

probability that excellence in medical care is being provided.

Historically, there has been acrimonious debate between frequentists and Bayesians. Frequentists have accused Bayesians of introducing an element of subjectivity into the analysis, via the prior distribution. Frequentists claim that their own analyses are objective, with parameter estimates being derived solely from the data. Bayesians respond that frequentists commonly use prior beliefs or knowledge in evaluating new data. Presented with results from a small clinical trial that found a null effect for a certain medication, one would interpret the results in light of previous large trials that found a significant and positive effect due to the medication. Bayesians claim that their analyses explicitly incorporate prior knowledge or beliefs, whereas frequentists implicitly incorporate prior knowledge in interpreting the results of their analyses. Similarly, Bayesians contend that choices in modelling strategies are often subjective. For instance, when modelling a binary outcome, clinical researchers frequently use a generalized linear model with a logit link (logistic regression). However, alternative, competing choices, such as the probit link or the complementary log–log link, are rarely considered, despite being viable alternatives.

The ability to incorporate prior evidence does not come without a risk. If one thinks that one is a perfect surgeon, and takes that as one's prior belief, it would take a lot of empirical evidence to shift one's perception. However, employing sensitivity analyses by varying the prior distribution allows one to examine the robustness of the results under different prior assumptions.

There is a growing interest in the use of Bayesian methods in clinical research. They are increasingly common in the area of diagnostic testing. Given that a test was positive, Bayes' theorem allows one to determine how the patient's prior probability or odds of having the condition have changed. Several commentators have argued that the *P*-value is of limited utility, is confounded with both effect size and sample size and tends to overstate the evidence against the null hypothesis (Freeman 1993; Lang *et al.* 1998; Goodman 1999b). There is a growing awareness of the potential usefulness of Bayesian methods in the

monitoring of, and analysis of, clinical trials (Berry 1993; Hughes 1993; Fayers *et al.* 1997). Bayesian methods allow an analysis of the strength of the trial's results to overcome different levels of scepticism about a treatment's effect. Bayesian methods allow more flexible interim monitoring of trials, and in assisting decisions to stop the trials. There is an increasing interest in the use of Bayesian methods in health care profiling (Christiansen & Morris 1997; Normand *et al.* 1997; Brophy & Joseph 1998). Bayesian hierarchical regression models allow profiling to be guided by medical standards, rather than by statistical standards. Bayesian methods are increasingly being recognized as a powerful tool in decision analysis (French & Smith 1997). In evaluating the results from a clinical trial, one must decide whether it is worse falsely to conclude that an ineffective treatment works, or falsely to conclude that an ineffective treatment does not work. If one can quantify the 'cost' associated with an incorrect decision, the costs can be incorporated directly into a Bayesian statistical analysis and combined with the data to produce an optimum decision (DeGroot 1970). Bayesian methods readily lend themselves to, and are increasingly being used in, meta-analysis (Stangl & Berry 2000).

In conclusion, the Bayesian paradigm allows one explicitly to incorporate prior knowledge and beliefs into statistical analyses. Bayesians view the data, as well as the underlying parameters, as random variables. Therefore, one is able to ask different questions of the data than are possible from the frequentist perspective. In particular, Bayesian methods allow one to construct intervals that have the interpretation that many scientists are tempted to ascribe to frequentist confidence intervals. Additionally, scientists are often tempted to interpret *P*-values as the probability that the null hypothesis is true – a statement that is nonsensical from a frequentist perspective, but perfectly natural to a Bayesian. Sensitivity analyses allow one to test the sensitivity of the results to the prior probability distribution that was specified. Bayesian methods allow the fitting of all the statistical models that one has become accustomed to fitting using frequentist methods. With the growth in the implementation of Bayesian methods in clinical research, clinicians – whether they subscribe to the Bayesian philosophy or not – need to be

aware of the existence and possibilities of these methods.

Acknowledgements

The authors thank Kathy Knowles Chapeskie for editorial assistance in preparing the manuscript. The views expressed herein are solely those of the authors and do not represent the views of any of the sponsoring organizations.

References

- Berry D.A. (1993) A case for Bayesianism in clinical trials. *Statistics in Medicine* **12**, 1377–1393.
- Brophy J.M. & Joseph L. (1998) Practice variations, chance and quality of care. *Canadian Medical Association Journal* **159**, 949–952.
- Christiansen C.L. & Morris C.N. (1997) Improving the statistical approach to health care provider profiling. *Annals of Internal Medicine* **127**, 764–768.
- Cox D.R. & Hinkley D.V. (1974) *Theoretical Statistics*. CRC Press, Boca Raton, FL.
- Davidoff F. (1999) Standing statistics right-side-up. *Annals of Internal Medicine* **130**, 1019–1021.
- DeGroot M.H. (1970) *Optimal Statistical Decisions*. McGraw-Hill, New York.
- Fayers P.M., Ashby D. & Parmar M.K. (1997) Tutorial in biostatistics: Bayesian data monitoring in clinical trials. *Statistics in Medicine* **16**, 1413–1430.
- Fraser D.A.S. (1976) *Probability and Statistics: Theory and Applications*. Institute for Theoretical Statistics, Toronto.
- Freedman L. (1996) Bayesian statistical methods [Editorial]. *British Medical Journal* **313**, 569–570.
- Freeman P.R. (1993) The role of *P*-values in analysing trial results. *Statistics in Medicine* **12**, 1443–1452.
- French S. & Smith J.Q. (1997) Bayesian analysis. In *The Practice of Bayesian Analysis* (eds S. French & J.Q. Smith), pp. 1–24. Arnold, London.
- Gelman A., Carlin J.B., Stern H.S. & Rubin D.B. (1995) *Bayesian Data Analysis*. Chapman & Hall, London.
- Gilks W.R., Richardson S. & Spiegelhalter D.J. (1996) Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo in Practice* (eds W.R. Gilks, S. Richardson & D.J. Spiegelhalter), pp. 1–19. Chapman & Hall, London.
- Goodman S.N. (1999a) Toward evidence-based medical statistics I: the *P*-value fallacy. *Annals of Internal Medicine* **130**, 995–1004.
- Goodman S.N. (1999b) Toward evidence-based medical statistics II: the Bayes factor. *Annals of Internal Medicine* **130**, 1005–1013.
- Hughes M.D. (1993) Reporting Bayesian analyses of clinical trials. *Statistics in Medicine* **12**, 1651–1663.
- Lang J.M., Rothman K.J. & Cann C.I. (1998) That confounded *P*-value. *Epidemiology* **9**, 7–8.
- Lee P.M. (1989) *Bayesian Statistics: an Introduction*, 2nd edn. Arnold, New York.
- Lilford R.J. & Braunholtz D. (1996) The statistical basis of public policy: a paradigm shift is overdue. *British Medical Journal* **313**, 603–607.
- Malakof D. (1999) Bayes offers a ‘new’ way to make sense of numbers. *Science* **286**, 1460–1464.
- Neyman J. (1950) *First Course in Probability and Statistics*. Henry Holt, New York.
- Neyman J. & Pearson E.S. (1933) On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society, Series A* **231**, 289–337.
- Normand S.L.T., Glickman M.E. & Gatsonis C.A. (1997) Statistical methods for profiling providers of medical care: issues and applications. *Journal of the American Statistical Association* **92**, 803–814.
- Royall R. (1997) *Statistical Evidence: a Likelihood Paradigm*. Chapman & Hall/CRC Press, Boca Raton, FL.
- Stangl D.K. & Berry D.A. (2000) *Meta-Analysis in Medicine and Health Policy*. Marcel Dekker Inc., New York.