

## IN BRIEF

- Three different interpretations of probability
- An understanding of the classical approach to hypothesis testing
- The underlying concepts of Bayesian statistical analysis
- Some applications of the Bayesian method
- A Bayesian approach to the evaluation of diagnostic and screening tests

## Further statistics in dentistry Part 9: Bayesian statistics

A. Petrie<sup>1</sup> J. S. Bulman<sup>2</sup> and J. F. Osborn<sup>3</sup>



**Statistics** can be defined as the *methods used to assimilate data, so that guidance can be given, and conclusions drawn, in situations which involve uncertainty*. In particular, **statistical inference** is concerned with drawing conclusions about particular aspects of a *population* when that population cannot be studied in full. Uncertainty arises here because the totality of the information is not available. Instead, to make inferences about the population, it is necessary to rely on a *sample* of data which is selected from the population; this sample data may be augmented, in certain circumstances, by auxiliary information which is obtained independently of the sample data. Clearly, uncertainty lies at the heart of statistics and statistical inference. This uncertainty is measured by a **probability** which therefore forms the crux of statistics and must be properly understood in order to interpret a statistical analysis.

### FURTHER STATISTICS IN DENTISTRY:

1. Research designs 1
2. Research designs 2
3. Clinical trials 1
4. Clinical trials 2
5. Diagnostic tests for oral conditions
6. Multiple linear regression
7. Repeated measures
8. Systematic reviews and meta-analyses
9. Bayesian statistics
10. Sherlock Holmes, evidence and evidence-based dentistry

### UNDERSTANDING STATISTICS AND PROBABILITY

#### Measuring probability

A probability is a number that takes some value equal to or between zero and one. If the probability of the 'event' of interest is zero, then the event *cannot* occur. So, for example, the probability of drawing an 'eleven' from a pack of cards is zero because there is no such card. If the probability of the event of interest is unity, then the event *must* occur. Most probabilities lie somewhere between the two extremes; the closer the probability is to one, the more likely the event, the closer it is to zero, the less likely the event.

#### Defining probability

To take a particular example, suppose it is of interest to determine the probability that a man has a DMFT of zero (the event of interest). What is really meant by 'probability' in this setting? There are various ways of understanding a probability, the three most common being based on the following interpretations:

1. **Frequency.** This view of probability, also called *frequentist* or *empirical* probability, forms the basis for what is termed the *frequentist* or *classical* approach to statistical inference. The probability is defined only in situations or 'experiments' which can (at least, theoretically) be repeated again and again in essentially the same circumstances, under the constraint that the result from any one experiment is independent of any other. Therefore, it cannot be applied to a 'one-off' event, such as assessing the probability that Prince Charles will be king. Strictly, although every event is one-off, many events can be regarded as similar enough to satisfy the criteria laid down by the frequentist approach. The frequency definition of probability is then the *proportion* of times the event of interest occurs when the 'experiment' is repeated on many occasions, and is equivalent to a *relative frequency*. The frequency definition of probability is easily understood in the context of coin tossing, when a single toss of the coin can be regarded as the experiment and obtaining a 'head' as the event of interest. If a fair coin were tossed 1,000,000 times (that is, a large number of times), the 'frequency' interpretation of the probability of a head would be the number of heads obtained divided by 1,000,000, ie the proportion of heads. In the DMFT example, if there are 1,000 men in the population, then each man is regarded as the experiment, and a DMFT of zero is regarded as the event of interest. The probability that a man has a DMFT of zero is the proportion of the 1,000 men with a DMFT of zero.
2. **Subjective** This view of probability, central to *Bayesian* inference, is also termed *personalistic* as it expresses the personal degree of

<sup>1</sup>Senior Lecturer in Statistics, Eastman Dental Institute for Oral Health Care Sciences, University College London;

<sup>2</sup>Honorary Reader in Dental Public Health, Eastman Dental Institute for Oral Health Care Sciences, University College London;

<sup>3</sup>Professor of Epidemiological Methods, University of Rome, La Sapienza

Correspondence to: Aviva Petrie, Senior Lecturer in Statistics, Biostatistics Unit, Eastman Dental Institute for Oral Health Care Sciences, University College London, 256 Gray's Inn Road, London WC1X 8LD  
E-mail: a.petrie@eastman.ucl.ac.uk

#### Refereed Paper

© British Dental Journal 2003; 194: 129–134

**Classical and Bayesian analyses**

The classical or Neyman–Pearson approach to statistical analysis relies on the frequency interpretation of probability whereas the Bayesian approach relies on a subjective interpretation of probability



belief an individual holds that an event will occur. For example, it may be an individual's personal view that a man from a particular population has a DMFT of zero, that a certain person has oral cancer, or that a coin will land on heads when tossed. The subjective view of probability is based on the individual's experiences and his or her ability to amass and construe information from external sources, and may well vary from one individual to another. It can be applied to one-off events.

3. **Model based.** This type of probability, sometimes termed an *a priori* probability, relies on being able to specify all possible equally likely outcomes of an experiment, in advance of or even without carrying out the experiment. So, in the coin tossing example, there are two equally likely outcomes, a head and a tail. The probability of an event which defines a particular outcome or set of outcomes, if appropriate, is the number of outcomes which relate to the event of interest divided by the total number of outcomes. Thus the probability of a head (the event of interest) is one divided by two (the total number of possible outcomes) which equals  $\frac{1}{2}$ . If a card were drawn from a pack of fifty two cards, the *a priori* probability of a 'heart' would be 13 (the number of hearts in the pack) divided by 52 (the number of cards), ie  $\frac{1}{4}$ . Clearly, only some situations are amenable to this approach to defining a probability; the probability of a man having a DMFT of zero cannot be assessed in this manner. It is interesting to note that the frequentist probability of an event tends to coincide with, or at least tends towards, the *a priori* probability when the experiment is repeated very many times. Thus, if a fair coin were tossed 10 times, it would not be surprising if 7 heads were obtained (giving a frequentist probability of a head as 0.7), but if the coin were tossed 10,000 times, the proportion of heads would be expected to be very close to 0.5.

**CONDITIONAL PROBABILITY**

There are various rules that can be adopted to evaluate probabilities of interest. Each will be illustrated by considering drawing a card or two cards from a pack.

1. **Addition rule.** This states that if two events are *mutually exclusive* (this means that if one of the events occurs, the other event cannot occur), then the probability that either one occurs is the sum of the individual probabilities. So for two events, A and B,

$$Pr(A \text{ or } B) = Pr(A) + Pr(B)$$

Thus the probability of drawing either a heart or a spade from the pack of 52 cards is  $\frac{1}{4} + \frac{1}{4} = \frac{1}{2} = 0.5$ .

2. **Multiplication rule.** This states that if two events are *independent* (this means that the events do not influence each other in any way), then the probability that both of these events occur is equal to the product of the probabilities of each. So,

$$Pr(A \text{ and } B) = Pr(A) \times Pr(B)$$

Thus the probability of drawing the king of hearts is  $\frac{1}{13} \times \frac{1}{4} = \frac{1}{52} = 0.019$ .

If the events are *not independent*, then a different rule, requiring the understanding of a *conditional probability*, has to be adopted. The conditional probability of an event B, written  $Pr(B|A)$  or  $Pr(B \text{ given } A)$ , defines the probability of B occurring when it is known that A has already occurred. The rule for dependent events states that the probability of both events occurring is equal to the probability of one times the *conditional probability* of the other. So,

$$Pr(A \text{ and } B) = Pr(A) \times Pr(B \text{ given } A)$$

For example, suppose that two cards are drawn from the pack and the first is not replaced before the second is taken. The probability of both of these cards being clubs is the product of the probability of the first being a club (ie  $\frac{13}{52}$ ) and the second being a club, given that the first was a club (ie  $\frac{12}{51}$ ), which is 0.085. Conditional probability plays an important role in Bayesian statistics.

**THE FREQUENTIST PHILOSOPHY**

The most common philosophy underlying statistical analysis is the **frequentist** or **classical** approach, often termed the **Neyman–Pearson** approach, named after the two statisticians who were instrumental in developing the early theory of statistical hypothesis tests. The two features which characterise the frequentist approach are:

1. All the information which is used to make inferences about the attributes of interest in the population is obtained from the sample.
2. The results of the analysis are interpreted in a framework which relates to the long-term behaviour of the experiment in assumed similar circumstances. Thus, the *P*-value, which is fundamental to the interpretation of the results, is strictly (although it is often misinterpreted) a *frequentist* probability.

Suppose a particular parameter, the population mean, is relevant to an investigation concerned with comparing the effects of a test and a control treatment on a response of interest. A sample of patients is selected and each patient is randomly allocated to one of the two treatments. For example, a double blind randomised trial (Fine *et al.*, 1985)<sup>1</sup> compared the mean wet plaque weight of adults' teeth (collecting the plaque from 20 teeth per adult) when one group of adults received an antiseptic mouthwash and

a second group received its vehicle control, each 'treatment' being used twice daily for nine months and in addition to normal tooth brushing. The null hypothesis,  $H_0$ , is that the means are the same in the two groups *in the population*. Using the *sample* data, a test statistic is evaluated from which a *P*-value is determined. The *P*-value is NOT the probability that the true difference in means is zero. Classicists regard the population attribute (in the above example, the difference in population means) as fixed, so that they cannot attach a probability directly to the attribute. The *P*-value is the probability of obtaining a difference between sample means equal to or more extreme than that observed, if  $H_0$  is true. That is, if the experiment were to be repeated many times, and  $H_0$  were true, the observed (or a more extreme) difference in means would be obtained on 100*P*% of occasions. In the same vein, the classicist strictly describes the 95% (say) confidence interval for the true difference in the two means as that interval which, if the experiment were to be repeated on many occasions, would contain the true difference in means on 95% of occasions.

It should be noted that the classical approach to hypothesis testing, because it does not allow a probability to be attached directly to the hypothesis,  $H_0$ , dichotomises the results according to whether or not they are 'significant', typically if the *P*-value is less or greater than 0.05. For this reason, it can be argued that the approach is not well suited to decision making, since the *P*-value does not give an indication of the extent to which  $H_0$  is false (eg how different the means are). If the sample size is large, the results of a test may be highly significant (ie the *P*-value very small) even if there is very little difference between the treatment means. Alternatively, the results may be non-significant (ie with a large *P*-value) if the sample size is small even if there is a large difference between the treatment means.

### THE BAYESIAN PHILOSOPHY

Bayesian statistical methods, developed from the reasoning adopted by an eighteenth century clergyman, the Rev. Thomas Bayes, draw a conclusion about a population parameter by combining information from the sample with initial beliefs about the parameter. More explicitly, the sample data, expressed as a **likelihood** function, is used to modify the **prior** information about the parameter, expressed as a probability distribution and derived from objective and/or subjective sources, to produce what is termed the **posterior** distribution for the parameter.

The Bayesian approach is described in detail in texts such as those by Iversen (1984)<sup>2</sup> and Barnett (1999),<sup>3</sup> and is summarised by Lilford and Brauholtz (1996).<sup>4</sup> It is characterised by the following features which differ quite markedly from those of the classical approach:

1. It incorporates information which is extraneous to the sample data into the calcula-

tions. This is the prior information about the parameter of interest.

2. It assumes that the parameter of interest, rather than being fixed, has a probability distribution. Initially, this is the prior distribution but it is updated, using the sample data, to form the posterior distribution. A probability distribution attaches a probability to every possible value of the quantity of interest. This means that, in a Bayesian analysis, it is possible to evaluate the probability that a parameter has a particular value and, consequently, the probability that a null hypothesis about the parameter is true. It is this probability in which most people are interested – the chance that null hypothesis is true – rather than the probability associated with the classical analysis, namely the *P*-value. Furthermore, a Bayesian can truly interpret a 95% confidence interval as the range of values which contains the true population parameter with 95% certainty, the interpretation often falsely adopted by classicists. The Bayesian point estimate of the parameter is usually taken to be the mode of the posterior distribution, ie its most likely value.
3. It relies on the subjective interpretation of a probability, reflecting a personal degree of belief in an outcome, as the choice of prior depends on the investigator and it is not interpreted in a frequentist manner. This personal belief in a parameter value or the truth of the null hypothesis will probably change as further evidence becomes available. The Bayesian accommodates this reasoning by using the sample data to update the prior into the posterior. Continual updating can be achieved by using this posterior as the prior for the next Bayesian analysis.

### The likelihood

Central to the Bayesian philosophy is the likelihood, the probability of getting the data observed in the sample when the parameter of interest takes a particular value (eg the value when  $H_0$  is true). The likelihood for the sample data will be different for different hypotheses about a particular parameter (or parameter specification such as the difference in means), ie for different values of this parameter. It is possible to consider all possible parameter values, and calculate the likelihood in each case, ie the probability of getting the data actually observed in each instance. This can be achieved if the parameter is assumed to follow a known probability distribution, such as the discrete Binomial or the continuous Normal distributions. If these probabilities are plotted against the parameter values, then the resulting plot is called the *likelihood function*.

### Bayes theorem

Bayes theorem provides the means of updating the prior probability using sample data, and is the basic tool of Bayesian analysis. Suppose a null hypothesis,  $H_0$ , about a particular parameter



### Bayes theorem

Bayes theorem provides a theoretical framework which enables an initial pre-experiment assessment of the probability of some event to be revised by combining the initial assessment with information obtained from experimental data

specification is to be tested, say that the difference in the mean responses between test and control treatments is zero in the population (eg that the difference between the mean plaque weights from adults' teeth after using either a mouthwash or a vehicle control for 9 months is zero). If the prior probability that  $H_0$  is true is  $Pr(H_0)$ , and the likelihood of getting the data when  $H_0$  is true is  $Pr(\text{data}|H_0)$ , where the vertical line is read as 'given', then Bayes theorem states that the posterior probability that  $H_0$  is true is:

$$Pr(H_0|\text{data}) \propto Pr(\text{data}|H_0) Pr(H_0),$$

ie the posterior probability is proportional to the product of the likelihood and the prior probability. Both the posterior probability and the likelihood are conditional probabilities. Bayes theorem converts the unconditional prior probability into a conditional posterior probability.

In fact, Bayes theorem states that:

$$Pr(H_0|\text{data}) = \frac{Pr(\text{data}|H_0)Pr(H_0)}{Pr(\text{data})}$$

where the denominator, called the *normalising* constant, is a factor which makes the total probability equal to one when all possible hypotheses are considered. When there are only discrete possibilities for the parameter values, say the set  $H_0, H_1, H_2, \dots, H_k$ , then the denominator becomes  $\sum Pr(\text{data}|H_i)Pr(H_i)$  where the sum extends over all possible values for  $i$ , namely  $i = 0, 1, 2, \dots, k$ .

When the parameter can take any value within a range of continuous values, then both the prior and posterior probabilities are replaced by probability densities, shown as smooth curves when plotted, with the area under each curve being unity (this corresponds to the sum of all probabilities being one). The posterior density function can then be used to evaluate the probability that the null hypothesis is true (ie that the parameter takes a particular value, say, zero) and also that the parameter takes values within a range such as between one and two.

#### The choice of prior

One of the factors which has limited the use of Bayesian analysis is the perception that its results are too dependent on an arbitrary factor, namely the choice of prior. In fact, where there is no information from the prior (it is a *non-informative prior* when all possible values for the parameter of interest are equally likely), the prior does not influence the posterior distribution, and the posterior and the likelihood will be proportional. At the other extreme, where the prior provides strong information (for example, when the prior suggests that there is only one possible value for the parameter), the likelihood will not influence the posterior and the posterior will be identical to the prior. Close to these two extremes, there are situations of *vague* and *substantial prior knowledge*. In the former case, the information in the sample data swamps the prior information so that the posterior and likelihood are virtually equal; in the latter case, the posteri-

or departs substantially from the likelihood. To further assuage those in doubt, it is possible to assess how robust conclusions are to changes in the prior distribution by performing a **sensitivity analysis**. Different priors, obtained perhaps from a number of clinicians, lead to a series of posterior distributions. In turn, these may or may not lead to different interpretations of the results, for example, about the extent to which it is believed a novel treatment may be beneficial when compared with an existing therapy.

There are a number of possible types of prior. These include:

1. **Clinical priors** – these express reasonable opinions held by individuals (perhaps clinicians who will participate in the trial) or derived from published material (such as a meta-analysis of similar studies).
2. **Reference priors** – such priors represent the weakest information (when all possible parameter values are equally likely, ie there is *prior ignorance*), and each is usually used as a baseline against which other priors can be compared.
3. **Sceptical priors** – these priors work on the basis that the effect of interest, such as the treatment effect measured by the difference in treatment means, is close to zero. In such situations, the investigator is sceptical about the effect of treatment, and wants to know the effect on the posterior of the worst plausible outcome.
4. **Enthusiastic priors** – these priors consider the spectrum of opinions which is diametrically opposed to that contained within the community of sceptical priors, namely when the investigator is optimistic about the treatment effect. He or she is interested in the effect on the posterior of the best plausible outcome.

It should be noted that sometimes, when the information extraneous to the sample data is limited, it is difficult or impossible to specify an appropriate prior. Then an **empirical Bayesian analysis**, in which the observed data is used to estimate the prior, may be performed instead of a full Bayesian analysis. Further details may be obtained from Louis (1991).<sup>5</sup>

#### APPLICATIONS OF THE BAYESIAN METHOD

Although the theory of Bayesian statistics has been around for many years, it has, in the past, been of limited application. This is because, usually, it is very difficult, if not impossible, to calculate the posterior distribution analytically. Instead, simulation techniques, such as Monte Carlo methods, have had to be used to approximate the distributions, and these are extremely computer intensive. However, with the advent of fast, cheap and powerful computers, and specialist software (such as: WinBUGS – [www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml](http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml)), this difficulty has, to a large extent, been overcome, and Bayesian methods are becoming more pop-

### Prior information



The experimental information concerning some event can be so overwhelming that it swamps the prior or pre-experiment information. Then the prior hardly influences the posterior or post-experiment probability of the event.



ular and finding a wider application. Some examples are discussed in the following subsections, more details of which can be obtained in papers such as those by Berry (1993),<sup>6</sup> Spiegelhalter *et al.* (1994),<sup>7</sup> Berry and Stangl (1996),<sup>8</sup> and Fayers *et al.* (1997).<sup>9</sup>

### Predictive probabilities

Since the Bayesian approach assumes a probability distribution for a parameter, it is possible to calculate **predictive probabilities** for the parameter values of *future* patients, given the results in the sample. This is impossible in the classical framework which is concerned with calculating the probability of the observed data, given a particular parameter specification. Thus, in the Bayesian framework, the potential exists to use the predictive probability, for example, to decide whether or not a specific future patient will respond to treatment, to predict the required drug dose for an individual, or to decide whether a clinical trial should continue.

### Diagnostic and screening tests

One of the easiest applications of the Bayesian approach, and one that was applied early on in the development of the method, is to the problem of diagnosis and screening (covered in an earlier paper, Diagnostic Tests for Oral Conditions, in this series). Although a dentist may rely on a formal test to diagnose a particular condition in a patient (oral cancer, say), it would be most unusual if the dentist does not have some preconceived idea of whether or not the patient is diseased. This subjective view might be based on the patient's clinical history and the presence of signs and symptoms (for example, a pre-cancerous lesion such as leucoplakia, erythroplakia, chronic mucocutaneous candidiasis, oral submucous fibrosis, syphilitic glossitis, or sideropenic dysphagia), or, if nothing is known about the patient, may simply be the prevalence of the condition in the population. It seems sensible to include such information in the diagnosis process, and this can be achieved fairly easily in a Bayesian framework. The preconceived idea is the prior (or pre-test) probability, the result of the diagnostic test (which may be positive or negative) determines the likelihood and Bayes theorem combines the two appropriately to produce the posterior (or post-test) probability that the patient has the condition. Rather than labouring through the mechanical process of applying Bayes theorem, a simple approach is to use Fagan's nomogram (Fig. 1).<sup>10</sup> The posterior probability is found by connecting the pre-test probability to the likelihood ratio, extending the line and noting where it cuts the post-test axis.

Consider the example which was used in Part 5 – Diagnostic Tests for Oral Conditions. A 17-month longitudinal study (Kingman *et al.*, 1988)<sup>11</sup> of 541 US adolescents initially aged 10–15 years was conducted with a view to using the child's baseline level of lactobacilli in saliva as a screening test for children at high risk of developing caries. A bacterial level of lactobacilli

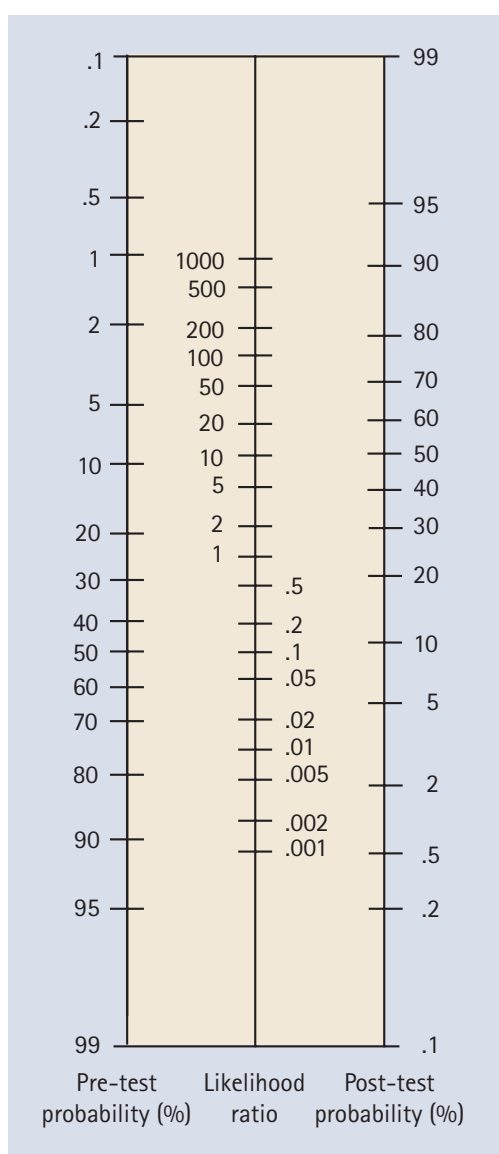


Fig. 1 Fagan's nomogram for interpreting a diagnostic test result. Adapted from Sachett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based Medicine: How to Practice and Teach EBM* (1997) by permission of the publisher Churchill Livingstone

>10<sup>5</sup> was regarded as a positive test result and this was compared with the child's caries increment after 17 months, where at least three lesions in the period were recorded as a positive disease result. Early detection of these high-risk children allows special preventative programmes to be instituted for them, and this is important both for the individual child and for society, as the gain can be expressed in terms of dental health and economy. Suppose that it is of interest to determine whether a particular child from the population under investigation is likely to be at high risk of developing caries. It is known that the prevalence of high risk children (in terms of caries development) in this population is about 21%, and the sensitivity and the specificity of the test are 15% and 93%, respectively. Thus the pre-test probability of the child being high risk can be taken as 0.21 (or 21%), and the likelihood ratio of a positive test result, which is the sensitivity divided by 100 minus the specificity (Petrie and Sabin, 2000),<sup>12</sup> is  $15/(100-93) = 2.14$ . Using Fagan's nomogram and connecting 21% on the left hand axis to 2.14 on the middle axis and extending the line

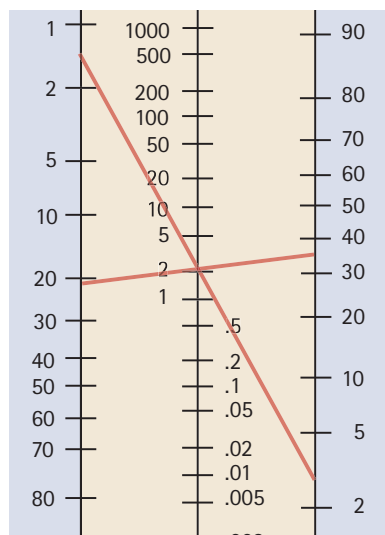


Fig. 2 Two examples of the use of Fagan's nomogram shown by the drawn red lines (a reduced section of Fig. 1 is shown)

gives a value on the right hand axis of about 35% so that the post-test or posterior probability is approximately equal to 0.35 (Fig. 2). On this basis, it is probably worth investigating the child further (eg by using additional test such as that based on the level of *mutans streptococci* in saliva). If, on the other hand, the child comes from a different population in which the prevalence of high risk children is only 1.5%, then the line connecting 1.5% on the left hand axis to 2.14 on the middle axis cuts the right hand axis at about 3% so that the post-test probability comes to approximately 0.03 (Fig. 2). The pre- and post-test probabilities are both extremely low in this instance, and the child from this population can be regarded as being at very low risk of developing caries, so that no further action needs be adopted for this child. It may be of interest to note that, in each case, the Bayesian post-test or posterior probability determined using Fagan's nomogram corresponds (after allowing for rounding errors and the approximations involved in the use of the nomogram) to the positive predictive value of the test, evaluated in the 'diagnostic tests' paper in this series.

### Clinical trials

The Bayesian approach can be used in clinical trials, both in their design and analysis, when a **decision** has to be made, such as whether or not to admit more patients to a study or to adopt a new therapy. The process requires an assessment of the costs and benefits of the consequences associated with the possible decisions. These consequences are expressed as utilities, and should be specified by an appropriate team of experts (eg dentists, pharmacologists, oncologists etc) who have to address issues relevant to the problem. These utilities are then weighted by the probabilities of the consequences (the predictive probabilities) to determine the expected

benefits. The decision that maximises the expected benefit or minimises the maximum loss is then chosen.

### CONCLUSION

The issue of whether or not to adopt a Bayesian approach to statistical analysis in a given circumstance remains controversial and one of personal choice. This paper has attempted to introduce the concepts and highlight the advantages and disadvantages of such procedures. Whatever one's views, however, it should be recognised that the Bayesian approach to data analysis is one of the greatest single developments in statistics since Pearson, Fisher, Gosset (Student) and their colleagues created the theoretical framework of the conventional significance test. After almost a century, and with the advent of powerful computers, the subject 'statistics' may be on the brink of a revolution as important as the change from Newton to Einstein was for Physics. In the early years of the twentieth century there were many sceptics about relativity and atomic physics just as now there are many sceptics among statisticians about the practical usefulness of Bayesian methods. However, the new millennium is still young and even if Bayesian methods are accepted universally, one wonders if the community of statisticians will be ready for a third revolution after another century!

- 1 Fine D H, Letizia J, Mandel I D. The effect of rinsing with Listerine antiseptic on the properties of developing dental plaque: *J Clin Periodontol* 1985; **12**: 660-666.
- 2 Iversen G R. Bayesian statistical inference. *Sage University Paper series on Quantitative Applications in the Social Sciences, series 07-043*. Thousand Oaks, California: Sage University Press, 1984.
- 3 Barnett V. *Comparative statistical inference*. 3rd edn. New York: Wiley, 1999.
- 4 Lilford R J, Braunholtz D A. The statistical basis of public policy: a paradigm shift is overdue. *Br Med J* 1996; **313**: 603-607.
- 5 Louis T A. Using empirical Bayes methods in biopharmaceutical research. *Stat Med* 1991; **10**: 811-829.
- 6 Berry D A. A case for Bayesianism in clinical trials. *Stat Med* 1993; **12**: 1377-1393.
- 7 Spiegelhalter D J, Freedman L S, Parmar M K B. Bayesian approaches to randomized trials. *J R Statist Soc A* 1994; **157**: 357-416.
- 8 Berry D A, Stangl D K. 'Bayesian methods in health-related research' in *Bayesian Biostatistics*. Eds Berry D A, Stangl D K. New York: Marcel Dekker, 1996.
- 9 Fayers P M, Ashby D, Parmar M K. Tutorial in biostatistics: Bayesian monitoring in clinical trials. *Stat Med* 1997; **16**: 1413-1430.
- 10 Sachett D L, Richardson N S, Rosenberg W, Haynes R B. *Evidence-based medicine: how to practice and teach EBM*. London: Churchill Livingstone, 1997.
- 11 Kingman A, Little W, Gomez I, Heifetz S B, Driscoll W S, Sheats R, Supan P. Salivary levels of *Streptococcus mutans* and lactobacilli and dental caries experiences in a US adolescent population. *Community Dent Oral Epidemiol* 1988; **16**: 98-103.
- 12 Petrie A, Sabin C. *Medical Statistics at a Glance*. Oxford: Blackwell Science, 2000.