

Clinical significance not statistical significance: a simple Bayesian alternative to p values

Paul R Burton, Lyle C Gurrin, Michael J Campbell

Abstract

Objectives—To take the common “Bayesian” interpretation of conventional confidence intervals to its logical conclusion, and hence to derive a simple, intuitive way to interpret the results of public health and clinical studies.

Design and setting—The theoretical basis and practicalities of the approach advocated is at first explained and then its use is illustrated by referring to the interpretation of a real historical cohort study. The study considered compared survival on haemodialysis (HD) with that on continuous ambulatory peritoneal dialysis (CAPD) in 389 patients dialysed for end stage renal disease in Leicestershire between 1974 and 1985. Careful interpretation of the study was essential. This was because although it had relatively low statistical power, it represented all of the data that were available at the time and it had to inform a critical clinical policy decision: whether or not to continue putting the majority of new patients onto CAPD.

Measurements and analysis—Conventional confidence intervals are often interpreted using subjective probability. For example, 95% confidence intervals are commonly understood to represent a range of values within which one may be 95% certain that the true value of whatever one is estimating really lies. Such an interpretation is fundamentally incorrect within the framework of conventional, frequency-based, statistics. However, it is valid as a statement of Bayesian posterior probability, provided that the prior distribution that represents pre-existing beliefs is uniform, which means flat, on the scale of the main outcome variable. This means that there is a limited equivalence between conventional and Bayesian statistics, which can be used to draw simple Bayesian style statistical inferences from a standard analysis. The advantage of such an approach is that it permits intuitive inferential statements to be made that cannot be made within a conventional framework and this can help to ensure that logical decisions are taken on the basis of study results. In the particular practical example described, this approach is applied in the context of an analysis based upon proportional hazards (Cox) regression.

Main results and conclusions—The approach proposed expresses conclusions in a manner that is believed to be a helpful

adjunct to more conventional inferential statements. It is of greatest value in those situations in which statistical significance may bear little relation to clinical significance and a conventional analysis using p values is liable to be misleading. Perhaps most importantly, this includes circumstances in which an important public health or clinical decision must be based upon a study that has unavoidably low statistical power. However, it is also useful in situations in which a decision must be based upon a large study that indicates that an effect that is highly statistically significant seems too small to be of practical relevance. In the illustrative example described, the approach helped in making a decision regarding the use of CAPD in Leicestershire during the latter half of the 1980s.

(J Epidemiol Community Health 1998;52:318-323)

Doctors are used to thinking in terms of probabilities based upon diagnostic tests; for example, the probability that someone has diabetes given that they have impaired glucose tolerance. It therefore seems quite natural to regard the statistical significance test as a form of diagnostic test and to try to use it to resolve research questions such as: “in workers exposed to asbestos, what is the probability that a diet rich in vitamin A reduces the risk of developing mesothelioma?” Unfortunately, questions such as this cannot properly be answered using significance tests within the framework of conventional statistics. As an alternative we advocate a simple Bayesian approach,¹ which can be implemented without specialist software and provides direct and understandable answers to questions of this type. A Bayesian approach to analysis can be invaluable when interpreting studies that must drive important decisions in clinical or public health medicine.^{1,2}

Medical statistics is founded upon the “frequency based” view of probability,³ that is, the long run frequency of an event in a series of identical experiments. Statistical inferences are based upon p values and confidence intervals.³ Unfortunately, p values are non-intuitive.¹ At first sight they appear to measure the probability that the null hypothesis is true (that is, the hypothesis that outcome is unaffected by an exposure variable of interest) and p values are commonly misinterpreted in this manner. However, the probability that an hypothesis is true or false is not the long run probability of an event and cannot even be expressed in the framework of frequency-based probability. In

Division of
Biostatistics and
Genetic Epidemiology,
TVW Telethon
Institute for Child
Health Research, PO
Box 855, West Perth
WA6872, Australia
P R Burton

Princess Margaret
Hospital for Children
and King Edward
Memorial Hospital,
Perth, Australia
L C Gurrin

Medical Statistics and
Computing, University
of Southampton,
Southampton General
Hospital,
Southampton, United
Kingdom
M J Campbell

Correspondence to:
Associate Professor
P Burton.

Accepted for publication
7 July 1997

any particular case, the hypothesis is either true or false and no frequency-based probability can, or should, be attached to it.³

Conventional statistical theory circumvents this difficulty by inverting the problem. Having considered every possible result that a study could potentially generate, a p value measures “the probability of observing data as extreme as, or more extreme than, the data actually observed assuming that the null hypothesis is true”.³ This is a statement of frequency-based probability, but the cost of remaining faithful to frequency-based probability is that the p value is often misunderstood. Significance at the 5% level is commonly interpreted to mean that “there is an effect (the null hypothesis is false)” while $p > 0.05$ is taken to mean that “there is no effect (the null hypothesis is true)”. These interpretations suggest that p values provide some direct quantification of the plausibility of the null hypothesis. However, a proper assessment of the plausibility of the null hypothesis requires the simultaneous consideration of the relative plausibility of other competing hypotheses. It cannot reasonably be based upon a single p value calculated assuming that the null hypothesis is true.

Degrees of belief and Bayesian statistics

The notion of probability can usefully be extended beyond its simple frequency-based interpretation. This leads to the concept of subjective probability, wherein probability is interpreted as representing a “degree of belief” about an unknown quantity or specified hypothesis.⁴ Sometimes such a statement will have no long run frequency interpretation. For example, you might hear that a political party has a “70% chance of winning the next election”. Clearly the election will not be repeatedly contested under identical conditions and a relative frequency interpretation of the probability does not exist. Nevertheless, it is still possible to place a sensible and objective interpretation on such probabilities, for example, by thinking of them in terms of betting odds.^{5,6}

Suppose you wish to conduct an observational or experimental study to learn about an unknown quantity or to gain evidence to support or refute a hypothesis. The Bayesian theory of statistical inference asks the researcher to incorporate pre-existing evidence and beliefs into what is called a prior distribution, which is simply a quantification of the current state of understanding about the unknown quantity. Values of the quantity that are viewed as having a high probability of being true are given a high prior probability while those that are viewed as being less likely are given a correspondingly lower prior probability. The choice of an appropriate prior distribution may entail the inspection of historical data, consultation with experts in the field or the development of theoretical models. Some statisticians have given considerable thought to formalising this process.⁷⁻⁹

Once the data collection for the study has been completed, the prior distribution is combined with the new information to produce a

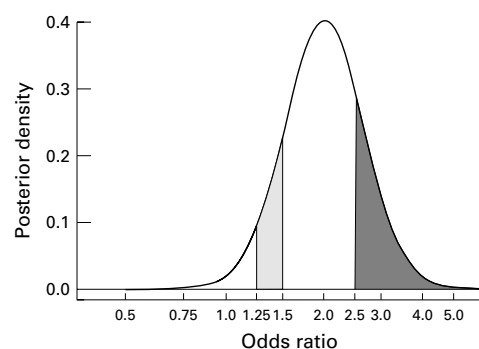


Figure 1 A Bayesian posterior distribution.

posterior distribution using a mathematical routine derived from Bayes’ Theorem.^{2,4,10} Conclusions are then based upon the posterior distribution. For example, the posterior probability that the true value of the quantity of interest lies between two values X and Y is equal to that proportion of the total area under the posterior distribution curve that falls between these two values. If the total area under the curve is normalised (scaled to equal one) the distribution is referred to as a density. The required posterior probability is then equal to the area under the density curve between X and Y. As an illustration, figure 1 details the posterior density for the odds ratio derived from a hypothetical case control study examining the association between childhood asthma and maternal smoking. The horizontal axis is on a logarithmic scale reflecting the fact that the fundamental scale of analysis is the scale of log (odds). The area under the density curve that falls between 1.25 and 1.5 (light shading) is 0.104, which means that, given whatever assumptions were made in choosing the prior distribution and given the observed data, there is a posterior probability of 10.4% that the true odds ratio lies somewhere between 1.25 and 1.5. Alternatively, the posterior probability that the true odds ratio is greater than or equal to 2.5 (dark shading) is 21.0%.

Bayesian inference supports direct statements about the probability of an hypothesis of interest or the magnitude of an unknown quantity and thus provides explicit answers to the types of question that are often asked in medical research. For example, the posterior probability that the hypothesis that “maternal smoking is associated with an increased risk of childhood asthma” is true would be equal to the area under the curve that falls above an odds ratio of 1.0. In the hypothetical example this would be 99.4%.

Objective Bayesian analysis and the uniform prior

The commentary above may reasonably be criticised for having glossed over the difficulties of choosing an appropriate prior distribution. Indeed, a common criticism of the Bayesian approach is that there are many realistic situations in which little or no pre-existing information is available and in choosing a prior distribution you are apparently trying to quantify

something that does not really exist. Alternatively, you may in some sense wish to let the data “speak for themselves”. In such a situation, some statisticians have proposed an objective Bayesian theory of statistical inference that uses “vague”, “flat” or “non-informative” prior distributions, which allows you to examine what the current data have to say about the hypothesis or quantity of interest without specific reference to prior beliefs.^{10 11} Such “non-informative” prior distributions often assign equal weight to all competing hypotheses and consider all possible values of unknown quantities to be “equally likely” before the new research is conducted. Such a prior may be referred to as being “uniform”¹⁰ on the scale of the main outcome variable. It is sometimes claimed that the uniform prior distribution provides a formal mathematical representation of a lack of prior information but this is untrue.^{5 6} In part this reflects the fact that uniformity is sensitive to transformation. For example, a prior distribution that is uniform on the scale of \log_e (odds) cannot simultaneously be uniform on the scale of proportions; an issue we will return to in the Discussion section. Nevertheless, it is true to say that the uniform prior focuses attention on current rather than pre-existing data⁹ and that it is a good “bench-mark”, which allows comparisons, from a “neutral” starting point, between data generated from different experiments or observational studies.

Estimating posterior probability using conventional confidence intervals

To remain faithful to the frequency based definition of probability, conventional 95% confidence intervals are properly defined in terms of hypothetical repetitions of a study and analysis: “if new studies were repeatedly carried out with the same analysis and a series of 95% confidence intervals calculated, 19 out of 20 such intervals would (in the long run) include the true value of the quantity being estimated”.³ However, most researchers interpret 95% confidence intervals rather differently. For example, if the 95% confidence intervals for an odds ratio extend from 1.6 to 6.4, many researchers would interpret this as indicating that there is a probability of 95% that the true odds ratio lies somewhere between 1.6 and 6.4.^{1 11} Formally, this interpretation is incorrect,³ because under the frequency-based view of probability the true odds ratio either does lie in this range or alternatively it does not; it has no long run interpretation. However, if an objective Bayesian analysis is carried out using a uniform prior distribution for the main outcome measure, conventional C% confidence intervals enclose a range of values that also encompass C% of the area under the posterior distribution.¹⁰ The 95% confidence intervals for an odds ratio that runs from 1.6 to 6.4 are therefore equivalent to a posterior probability of 95% that the true odds ratio lies between 1.6 and 6.4, which is analogous to the common interpretation of the confidence intervals stated above. Although this should formally be referred to as a Bayesian 95% credible

KEY POINTS

- Decision making in public health and clinical medicine is sometimes seriously impaired by the misinterpretation of p values and confidence intervals.
- Bayesian-based inferences are more intuitive and can lead to more appropriate decisions.
- Given a uniform prior distribution conventional C% confidence intervals cover C% of the Bayesian posterior distribution.
- This equivalence leads to a simple way to generate useful Bayesian inferences from a conventional analysis using standard software.
- This approach to inference can be very helpful for decision makers in public health and clinical medicine.

interval¹² (a range of values encompassing 95% of the posterior probability) this interpretation of 95% confidence intervals is acceptable provided that you recognise that it is based upon subjective probability, not frequency probability, and that any prior information is viewed as being “vague”.

It is the congruence between conventional confidence intervals and Bayesian posterior probability—when the prior distribution is uniform—which forms the basis of a simple approach to statistical inference that we have recently advocated.¹ This article illustrates the proposed approach with an example, which we hope makes it easy to understand and straightforward for public health and clinical researchers to implement.

A practical example: comparing mortality on continuous ambulatory peritoneal dialysis and haemodialysis

In 1976, continuous ambulatory peritoneal dialysis (CAPD) was introduced as a new treatment for end stage renal failure.¹³ It revolutionised the provision of renal replacement therapy, particularly in the United Kingdom where haemodialysis (HD) facilities had historically been very limited. In Leicestershire the acceptance rate for new patients more than doubled between 1976 and 1985.¹⁴ However, by the mid 1980s evidence had accumulated that suggested that the risk of death on CAPD might be greater than that on HD.¹⁴⁻¹⁶ In consequence, the Director of Renal Services in Leicestershire became concerned that the policy of putting most new patients onto CAPD might be unwise. A multi-centre prospective cohort study had been set up in the United Kingdom,¹⁷ but it was not due to produce definitive results for several years. It was therefore decided that a historical cohort study should be undertaken in Leicestershire itself, to determine whether a rational short-term policy decision could be taken regarding the use of CAPD while waiting for the larger study to report.

A retrospective case notes review was undertaken of patients who received renal

replacement therapy in Leicester between June 1974 and July 1985. Information was collected pertaining to the date of commencement of treatment, the dates of any transfers between therapies, and, where relevant, the date of death. A detailed record was also made of potential confounding factors present at the start of treatment. Twenty two deaths were observed in 177.8 patient years on CAPD and 60 deaths in 822.7 patient years on HD. Given the need to model irregular treatment changes and to adjust for confounding variables, survival analysis was undertaken using Cox's proportional hazards regression.^{18, 19}

Having adjusted for relevant confounders, the estimated relative risk of death (CAPD:HD) was 0.77 (95% confidence intervals 0.44 to 1.35); that is, the adjusted risk of death on CAPD was lower than that on HD.^{14, 20} In a standard test against the null hypothesis (H_0 : relative risk = 1.0) the two tailed p value was 0.36. It was initially concluded that because the relevant p value was non-significant, the null hypothesis should be accepted and the inference drawn that the risk of death on CAPD was the same as that on HD. However, this conclusion was not justified; the significance level of a test against the null hypothesis should never be used to conclude that the null hypothesis is true.³ The null hypothesis was merely one of a range of different hypotheses that were all consistent with the data; 95% confidence intervals indicate that the risk of death on CAPD may have been higher, lower or the same as that on HD. This problem arose because, although the study used all of the available data, there were only 82 deaths in total and the analysis lacked power. It might therefore have been tempting to infer that the study was too small to draw a meaningful conclusion. However, this would have failed to use the available data to their maximum potential and future patients may then have received less than ideal treatment. As an alternative, a simple Bayesian analysis invoking a uniform prior distribution was performed.^{14, 20}

The main outcome of the Cox model is the \log_e (relative risk (RR)). Conventional confidence intervals on this scale are symmetrical about the estimated \log_e (RR) and we will use a prior distribution that is uniform on this scale. The estimated \log_e (RR) is -0.261 and its standard error is 0.286 . Consider the confidence intervals with an upper limit at \log_e (RR)= 0.0 . By symmetry around the estimate (-0.261), the corresponding lower limit is -0.522 . The relevant confidence intervals therefore extend $0.261/0.286 = 0.913$ standard errors on either side of the estimate. Reference to a table of normal values shows that $Z=0.913$ equates to about a 64% confidence interval. Given the congruence between Bayesian and conventional analyses when the prior distribution is uniform, the posterior probability that the true \log_e (RR) is less than zero ($RR < 1.0$) is therefore equal to the probability that it falls within the confidence intervals (64%) plus the probability that it falls below the confidence interval ($18\% = (100\% - 64\%)/2$). The posterior probability that the true mortality on

CAPD is lower than that on HD is therefore 82%. Equivalently there is an 18% posterior probability that the risk of death on CAPD is higher than that on HD. This is equal to the one tailed p value,^{11, 21} which is half the two tailed p value of 0.36.

In addition, the Bayesian approach allows you to go beyond what is feasible in a conventional analysis. As an example, you may choose to estimate the probability that the mortality on CAPD is higher than that on HD to an extent that might seriously impair its clinical value; for example, the probability that the relative risk (CAPD:HD) is greater than 1.25. On the scale of \log_e (RR) the upper limit of the relevant confidence intervals is now $\log_e(1.25) = 0.223$. By symmetry the corresponding lower limit is -0.745 ($RR=0.48$). This equates to 1.69 standard errors on either side of the estimate and a 91% confidence interval. Assuming a uniform prior, the posterior probability that the relative risk is less than 1.25 is therefore $91\% + 4.5\% = 95.5\%$ and there is only a 4.5% posterior probability that it is greater than 1.25.

These results were interpreted as indicating that the probability that the risk of CAPD was greater than that on HD to an extent that would seriously impair its clinical usefulness was small. It was therefore concluded that it was both safe and appropriate to continue treating most new patients in Leicestershire with CAPD. This conclusion was later supported by the results of the multi-centre prospective cohort study.¹⁷ Thus, despite the limited size and low power of the study, a simple Bayesian analysis invoking a uniform prior permitted a rational evidence-based treatment policy to be determined for renal replacement therapy in Leicestershire for the mid 1980s.

Discussion

The theoretical basis of the approach we propose is not new.¹⁰ The Bayesian interpretation of conventional confidence intervals is explicitly²²⁻²⁸ or implicitly²⁹⁻³² outlined in a number of well known medical statistics texts. As far back as 1961, Birnbaum,³³ writing in a statistics journal, considered the use of curves through a range of paired confidence limits (at different levels of confidence) as an "omnibus technique for estimation and testing statistical hypotheses". More recently, Pocock and Hughes have argued³⁴ that, in certain circumstance, 70% confidence intervals that stretch just over one standard error from the sample mean may usefully be used to complement conventional 95% confidence intervals because it may provide "valuable supplementary information on the uncertainty about the point estimate". Furthermore, in recommending a related approach for interpreting clinical trials, Hughes¹¹ notes that "(its) use is unlikely to be controversial amongst statisticians whilst being intuitively appealing to clinicians". Nevertheless, despite its extensive theoretical underpinning, the approach we describe seems unknown to most researchers in public health and clinical science and it is not properly described in any publication that is easily accessible to scientists without statistical training.

In a thought provoking article in the *British Medical Journal*, Lilford *et al* advocated the increased use of Bayesian methods emphasising that they may be used “to calculate probabilities that may be extrapolated directly to clinical practice”.³⁵ In a later article, Lilford and Braunholtz highlighted the potential value of Bayesian inference in the decision making process in public policy.² Likewise, in an earlier paper in *Statistical Science*,³⁶ Breslow stated that Bayesian methods “offer a natural setting for synthesis of expert opinion in deciding policy matters” and “it seems clear that Bayesian and frequentist approaches each will have a role to play in biostatistical applications in years to come”. We fully endorse all of these views. However, we believe that the introduction of Bayesian methods continues to be hindered by the historical difficulties of conducting such analyses. Despite the interest generated by Lilford and Braunholtz’s article,² it was never intended as a “how to do it” guide and, although appetites may have been whetted, very few readers would know where to start should they wish to conduct an analysis of their own. However, all this may soon change. With the arrival of general purpose Bayesian software such as BUGS³⁷ (Bayesian inference Using Gibbs Sampling) and First Bayes,³⁸ it is probable that Bayesian analyses will soon become common place. The time is therefore ripe to start exploring Bayesian methods. The approach we advocate represents one of the simplest types of Bayesian analysis and enjoys the particular advantage that it can be based upon the results of a conventional analysis and can therefore be undertaken using standard software. We believe that the approach is not only of value in its own right, but that many will find it to be a useful stepping stone to the use of more sophisticated Bayesian methods in the future.

Our approach is most useful in situations in which a conventional analysis may be difficult or misleading. These include circumstances in which: (a) a statistically non-significant result is large enough to be clinically relevant (small sample size); (b) a statistically significant result is too small to be of clinical relevance (very large sample size); or (c) it is desired that conclusions are drawn about the probable similarity of two outcomes without concluding that non-significant means that there is no difference. Although the approach should not be used as an excuse for designing studies that are too small we believe that it is a useful adjunct to conventional methods in medical statistics.

The approach is explicitly dependent upon the use of a uniform prior distribution (technically a prior that is jointly uniform for all quantities being estimated). Some Bayesians argue that it is invalidated by the fact that no attempt is made to formally quantify pre-existing information in a subjective prior distribution. We acknowledge that neither the proposed approach nor a traditional analysis takes proper account of prior information; indeed this is sometimes claimed to be a strength of conventional statistics. If there is important pre-existing information that needs to be taken into

account, it must be incorporated formally in the analysis (some form of meta-analysis) or informally in the qualitative process of drawing conclusions. In this sense our approach is no different to a conventional analysis. Furthermore, Lindley¹⁰ argues that it may be appropriate to use the uniform prior distribution even when there is pre-existing information because: “even when one has some appreciable prior knowledge of theta (a quantity to be estimated) one may like to express the posterior beliefs about theta without reference to them (the prior knowledge)”. Spiegelhalter³⁷ comments: “The appropriate specification of ‘non-informative’ priors is an old problem in Bayesian statistics, and is particularly important when techniques are to be used in a scientific context in which an ‘objective’ inference is required”.

An important difficulty is that a prior distribution that is uniform on one scale cannot be uniform on an alternative scale. If two scales of analysis are equally appropriate, and the use of a prior that is uniform on one scale leads to a qualitatively different conclusion to an analysis based upon a prior that is uniform on another, these conclusions must be viewed as uncertain. Fortunately, like others,¹¹ we have found that unless the sample size is very small, alterations of scale generally make small quantitative changes to the estimates of interest rather than large qualitative differences to the principal conclusions. If a researcher feels strongly that all reasonable values of a quantity of interest are not equally likely a priori, for example if he or she has reason to believe that the null hypothesis is particularly likely to be true, then it would be misleading to use the approach we describe. However, it would then be equally misleading to interpret confidence intervals in the Bayesian manner, which is so commonly advocated and used (see above). Logically, in any setting in which it is reasonable to interpret 95% confidence intervals as the range of values within which you are 95% certain that the true value of whatever you are estimating really lies, then it is equally reasonable to use the methods we advocate. The sensitivity of conclusions to a particular choice of prior distribution will, in general, diminish as the sample size of the current data set increases. Nevertheless, inferences may be sensitive to the choice of prior in very small data sets and under such circumstances a sensitivity analysis using a range of priors would be desirable.

Many medical researchers already interpret conventional confidence intervals as if they were Bayesian credible intervals.¹¹ What we are proposing does not therefore entail a radical change in the way researchers naturally think about their data. Nevertheless, it might help to encourage a wider recognition that this common interpretation of confidence intervals is only valid if you work within a Bayesian framework using a uniform prior distribution. Given that it is valid in this framework and given that such an approach is potentially so informative, we would argue that it may sometimes be useful to take it to its logical conclusion by constructing and then

interpreting the full posterior distribution generated from a Bayesian analysis using a uniform prior. The inferences generated in this way are more informative and more easily understood than conventional p values and confidence intervals and they provide direct answers to many of the important types of question that are commonly asked by researchers in clinical and public health medicine.

We would like to thank Drs Jennifer Kurinczuk and Max Parmar for helpful comments on earlier drafts. We would also like to thank two anonymous reviewers for constructive advice. The original research that investigated the mortality and morbidity of patients on renal replacement therapy was conducted in association with Professor John Walls and was supported by the Leicester Renal Research Fund and the Medical Research Council.

Funding: this study was funded by the National Health and Medical Research Council of Australia as one component of Program Grant #96/3209.

- 1 Burton PR. Helping doctors to draw appropriate inferences from the analysis of medical studies. *Stat Med* 1994;**13**:1699–713.
- 2 Lilford RJ, Braunholtz D. The statistical basis of public policy: a paradigm shift is overdue. *BMJ* 1996;**313**:603–7.
- 3 Armitage P, Berry G. *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell Scientific, 1994: 76, 77, 94–99.
- 4 Lindley DV. *Introduction to probability and statistics from a Bayesian viewpoint. Part 1 Probability*. Cambridge: Cambridge University Press, 1965: 19–25, 29–42, 50, 58.
- 5 Walley P. *Statistical reasoning with imprecise probabilities*. London: Chapman and Hall, 1991.
- 6 Walley P, Gurrin L, Burton P. Analysis of clinical data using imprecise prior probabilities. *The Statistician* 1996;**45**:457–86.
- 7 Freedman LS, Spiegelhalter DJ. The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *The Statistician* 1983;**32**:153–60.
- 8 Spiegelhalter DJ, Freedman LS. Bayesian approaches to clinical trials. In: Bernardo JM, DeGroot MH, Lindley DV, et al, eds. *Bayesian statistics 3*. Amsterdam: North-Holland, 1988.
- 9 Chaloner K, Church T, Louis TA, Matts JP. Graphical elicitation of a prior distribution for a clinical trial. *The Statistician* 1993;**42**:341–53.
- 10 Lindley DV. *Introduction to probability and statistics from a Bayesian viewpoint. Part 2 Inference*. Cambridge: Cambridge University Press, 1965: 1–13, 15, 18, 19.
- 11 Hughes MD. Reporting Bayesian analyses of clinical trials. *Stat Med* 1993;**12**:1651–63.
- 12 Winkler RL. *An introduction to Bayesian inference and decision*. New York: Holt, Rinehart and Winston, 1972: 395–6.
- 13 Popovich RP, Moncrief JW, Dechedr JB, et al. The definition of a novel portable/wearable equilibrium peritoneal dialysis technique. [Abstract]. *Abstracts of the American Society of Artificial and Internal Organs* 1976;**5**:64.
- 14 Burton PR. *Modelling survival on renal replacement therapy*. Leicester: University of Leicester, 1992. MD Thesis 1.31.1.45–1.53, 3.39–3.42.
- 15 Kjellstrand CM. Introduction to a workshop on morbidity and mortality in hemodialysis, hemofiltration and continuous ambulatory peritoneal dialysis. *American Society of Artificial and Internal Organs* 1983;**6**:167–8.
- 16 Kramer P, Broyer M, Brunner FP, et al. Combined report on regular dialysis and transplantation in Europe, XIV, 1983. *Proceedings of the European Dialysis and Transplantation Association* 1984;**21**:5–68.
- 17 Gokal R, Jakubowski C, King J, et al. Outcome in patients on continuous ambulatory peritoneal dialysis and haemodialysis: 4-year analysis of a prospective multi-centre study. *Lancet* 1987;**ii**:1105–9.
- 18 Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society (series B)* 1972;**34**:187–220.
- 19 BMDP Statistical Software. *User manual*. Berkeley: University of California Press, 1981.
- 20 Burton PR, Walls J. Selection-adjusted comparison of life-expectancy of patients on continuous ambulatory peritoneal dialysis, haemodialysis and renal transplantation. *Lancet* 1987;**ii**:1115–19.
- 21 Greenland S. P values, hypothesis tests and likelihood: implications for epidemiology of a neglected historical debate. [Letter]. *Am J Epidemiol* 1994;**139**:116–17.
- 22 Armore SJ. *Introduction to statistical analysis and inference for psychology and education*. New York: John Wiley, 1966: 308.
- 23 Kirkwood BR. *Essentials of medical statistics*. Oxford: Blackwell Scientific, 1988: 27.
- 24 Matthews DE, Farewell VT. *Using and understanding medical statistics*. Basel: Karger, 1988: 98.
- 25 Mould RF. *Introductory medical statistics*. Bristol, IOP Publishing, 1989: 48.
- 26 Duncan RC, Knapp RG, Miller MC. *Introductory biostatistics for the health sciences*. 2nd ed. New York: John Wiley, 1983: 70.
- 27 Bland M. *An introduction to medical statistics*. Oxford: Oxford University Press, 1987: 139.
- 28 Campbell MJ, Machin D. *Medical statistics: A commonsense approach*. 2nd ed. Chichester: John Wiley, 1993: 64.
- 29 Colton T. *Statistics in medicine*. Boston: Little, Brown, 1974: 127.
- 30 Petrie A. *Lecture notes on medical statistics*. Oxford: Blackwell Scientific, 1978: 47.
- 31 Bradford Hill A. *A short textbook of medical statistics*. 11th ed. London: Hodder and Stoughton, 1984: 107.
- 32 Rothman KJ. *Modern epidemiology*. Boston: Little, Brown, 1986: 119, 122.
- 33 Birnbaum A. Confidence curves: an omnibus technique for estimation and testing statistical hypotheses. *Journal of the American Statistical Association* 1961;**56**:246–9.
- 34 Pocock SJ, Hughes MD. Estimation issues in clinical trials and overviews. *Stat Med* 1990;**9**:657–71.
- 35 Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. *BMJ* 1995;**311**: 1621–5.
- 36 Breslow N. Biostatistics and Bayes. *Statistical Science* 1990;**5**:269–98.
- 37 Spiegelhalter D, Thomas A, Best N, Gilks W. *BUGS Bayesian inference Using Gibbs Sampling version 0.50*. Cambridge: MRC Biostatistics Unit, 1995. (See world wide web page <http://www.mrc-bsu.cam.ac.uk>).
- 38 O'Hagan A. *Instructions for using First Bayes version 1.0*. Nottingham: University of Nottingham, 1994.