STATISTICS IN MEDICINE, VOL. 12, 1651–1663 (1993)

REPORTING BAYESIAN ANALYSES OF CLINICAL TRIALS

MICHAEL D. HUGHES

Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston MA 02115, U.S.A.

SUMMARY

Many clinicians wrongly interpret p-values as probabilities that treatment has an adverse effect and confidence intervals as probability intervals. Such inferences can be validly drawn from Bayesian analyses of trial results. These analyses use the data to update the prior (or pre-trial) beliefs to give posterior (or post-trial) beliefs about the magnitude of a treatment effect. However, for these methods to gain acceptance in the medical literature, understanding between statisticians and clinicians of the issues involved in choosing appropriate prior distributions for trial reporting needs to be reached. I focus on two types of prior that deserve consideration. The first is the non-informative prior giving standardized likelihood distributions as post-trial probability distributions. Their use is unlikely to be controversial among statisticians whilst being intuitively appealing to clinicians. The second type of prior has a spike of probability mass at the point of no treatment effect. Varying the magnitude of the spike illustrates the sensitivity of the conclusions drawn to the degree of prior scepticism in a treatment effect. With both, graphical displays provide clinical readers with the opportunity to explore the results more fully. An example of how a clinical trial might be reported in the medical literature using these methods is given.

1. INTRODUCTION

Recent developments in the application of Bayesian methods to the design and analysis of clinical trials have been reviewed by Spiegelhalter and Freedman.[1] Essentially, Bayesian methods use data from a clinical trial to update prior or pre-trial beliefs about a treatment's effect to give posterior or post-trial beliefs. This allows an analysis of the strength of the trial's results to overcome different prior beliefs. Whilst Bayesian methods have been used in such sensitivity analyses for internal decision making within a pharmaceutical company,[2] they have been used rarely in reporting results to medical audiences or to drug regulatory authorities. However, the direct interpretation of probability statements about the magnitude of a treatment's effect and the simplicity of understanding portrayed by graphical displays of posterior distributions should encourage wider use of these methods.

In this paper the role of Bayesian analyses in the reporting of clinical trials is considered with an emphasis on the use of two types of prior belief as aids for interpreting results. Two phase III clinical trials, both having a binary (survival) endpoint, are used as illustrations. A Bayesian approach to their analysis is described in the next section. In Section 3 a non-informative prior distribution is considered, giving an analysis which is essentially data-dependent and so unlikely to be controversial. In Section 4 a prior distribution that might represent the beliefs of an individual who is sceptical of the existence of a treatment effect is considered. By varying the degree of scepticism, the ability of the trial data to overcome weak or strong prior opinions against the treatment can be examined. Section 5 gives an example of a report using these approaches in a manner suitable for publication in the medical literature. The example also

0277–6715/93/181651–13$11.50

Received February 1991
Revised December 1992

illustrates the use of prior distributions based on the trial investigators' pre-trial beliefs and on results from previous similar trials. In Section 6, I discuss the use of these methods.

2. BAYESIAN ANALYSES FOR THE BAYREP AND AIMS TRIALS

Two trials with somewhat contrasting results have been chosen to illustrate how Bayesian analyses might be presented. The first, known as the Anistreplase Intervention Mortality Study (AIMS), is a placebo-controlled trial designed to assess whether a thrombolytic agent, anistreplase, reduces all-cause mortality within one year of an acute myocardial infarction. This trial has been reported in the medical literature[3] using classical (frequentist) methods of analysis. Of 634 patients randomized to receive placebo, 133 (17·8 per cent) died within one year, compared with 69 of 624 (11·1 per cent) randomized to receive anistreplase, giving an odds ratio of 0·57.

The second trial, used in the example in Section 5, is not yet completed and so no real data can be presented; I refer to it as the BAYREP trial. It is a placebo-controlled trial designed to assess whether a beta-blocker reduces all-cause mortality during two years of treatment for patients with oesophageal varices that have not previously bled. This trial is of particular interest as the investigators involved have described their prior beliefs for the treatment effect, and also the magnitude of effect needed before they would use beta-blockers in their routine clinical practice. The 'data' used in the example are based on the result of a similar, but smaller, trial reported in a conference abstract,[4] but with the counts increased by a factor of 5 to reflect the planned size of the BAYREP trial.

As both trials have mortality as their primary endpoint, a simple analysis of the data might use a logistic model for the probability π of dying during follow-up:

$$\ln\frac{\pi}{(1-\pi)} = \alpha + \beta x$$

where $x = 0$ if a patient receives placebo and $x = 1$ if the active treatment. Here β is the log odds ratio of dying for the treatment relative to placebo, and α (which is of less interest) is the log odds of dying in the placebo group. I assume a uniform non-informative prior distribution for α which is independent of β. On a log odds scale, this gives all values of α equal probability *a priori*, and so $p(\alpha)$ is an (improper) uniform distribution. For simplicity, the analyses presented describe the marginal post-trial density function for β given the trial's data y, obtained using Bayes' theorem:

$$p(\beta|y) = k\int l(y|\alpha, \beta)p(\beta)\mathrm{d}\alpha$$

where $l(y|\alpha, \beta)$ is the likelihood function based on the logistic model, $p(\beta)$ is the prior distribution for the treatment effect and k is a constant chosen so that the $p(\beta|y)$ integrates to unity over the range of β. Each of the next two sections describes analyses based on a different choice of the distribution, $p(\beta)$.

3. THE NON-INFORMATIVE PRIOR AND STANDARDIZED LIKELIHOODS

Many doctors interpret classical p-values and confidence intervals in a Bayesian manner. For instance, a p-value is the probability of obtaining, under the null hypothesis (usually of no treatment effect), an effect as or more extreme than that observed. However, when the effect observed is beneficial, the p-value is often interpreted as the probability that treatment causes a detrimental effect. A similar misinterpretation of confidence intervals as probability intervals is

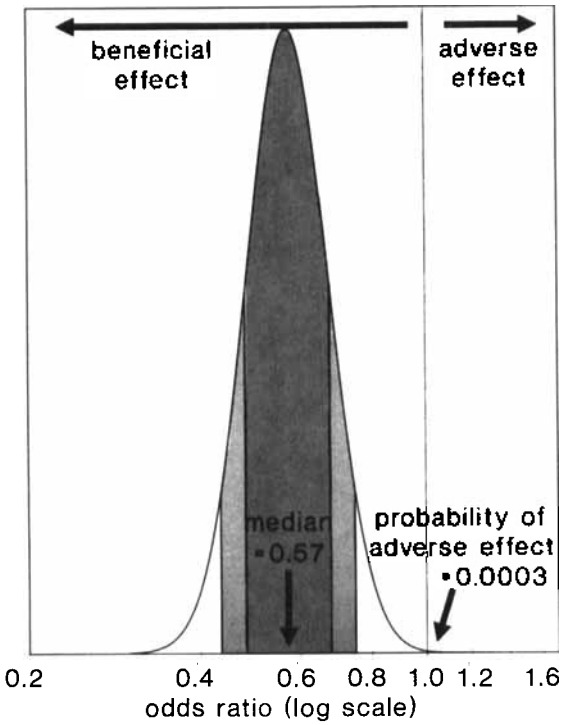

Figure 1. The standardized likelihood distribution for an effect of anistreplase relative to placebo in the AIMS trial

also commonly made. The recent proposal by Pocock and Hughes[5] that 70 and 95 per cent confidence intervals be overlaid to indicate more strongly the likely size of a treatment effect leans towards a Bayesian probability interpretation of these intervals.

Using the non-informative prior for β in which $p(\beta)$ is an (improper) uniform distribution leads to the posterior distributions known as the standardized likelihood distribution:

$$p(\beta|y) = k\int l(y|\alpha, \beta)\mathrm{d}\alpha$$

Figure 1 shows this distribution for the AIMS trial. The median treatment effect shown on the plot is very close to the expected treatment effect as the likelihood function is approximately symmetric. Also shown are probability intervals: overlaying a 70 per cent interval (heavier shading) on a 90 per cent interval (lighter shading) emphasizes the range of values for which the standardized likelihood distribution is most concentrated and shows clearly the degree of uncertainty still prevailing about the likely size of the treatment effect. Each interval is constructed by finding the two limits which exclude equal tail probabilities: for example, 0·05 for the 90 per cent interval. Alternatively, the highest posterior density intervals (the shortest intervals containing the required probability) might be used. For unimodal symmetric likelihood functions each approach gives the same interval. With skew distributions the former might be preferred because their interpretation is simpler, there being a clear relationship between an interval bounded by the point of no effect and the probability of a detrimental effect.

Also shown on Figure 1 is the probability of an adverse effect of treatment (an odds ratio greater than one). With large samples, the standardized likelihood distribution will approximate a normal distribution and so these probabilities will be very similar in magnitude to the p-values

obtained from a classical one-tailed significance test but need doubling to give the p-value corresponding to a two-tailed test. Similarly, a confidence interval with coverage $(1 - p)$ will approximate a $(1 - p)$ probability interval. Thus the presentation of statistical results in terms of characteristics of the standardized likelihood distribution allows clinically relevant interpretations to be drawn whilst maintaining a link with the classical interpretation.

Another useful summary of the possibility of an adverse effect is the ratio of the probability of a beneficial effect to an adverse effect: for the AIMS trial this is 0·9997/0·0003 = 3332:1. Of course, the point of no effect may be of less practical interest than other points if, for instance, costs or side effects of treatment mean that a greater treatment effect is necessary to justify its routine use. Providing a plot of the standardized likelihood distribution enables readers of the trial report to assess the evidence for an effect that they judge to be clinically worthwhile. This is illustrated in the example in Section 5. In addition, for trials of adequate size the analysis might be extended to show any dependence on patient characteristics of the ratio of probabilities for a beneficial effect to an adverse effect.

4. SCEPTICAL PRIOR BELIEFS

Clinicians have their own beliefs about the possibility that a treatment will have a beneficial effect. These beliefs may be based on results from other trials or on their own experiences with the treatment or related treatments, or they may be more subjective. If a treatment is to become established in clinical practice, then data from a trial need to affect such beliefs so that clinicians are convinced of its worth. In this section, one conceptually simple form for the beliefs of a sceptical clinician is considered. Analysis leads to graphical displays showing how the post-trial belief in a beneficial effect of a treatment is affected by the degree of prior belief in no treatment effect.

To introduce this prior distributions, consider the prior belief of one particular investigator prior to starting the BAYREP trial (Figure 2). This prior distribution was elicited (as for other investigators) in a postal questionnaire using the framework suggested by Spiegelhalter and Freedman[6] and developed by Gore.[7] Note the spike at the point of no effect, around which there is a fairly uniform distribution. Discussion with this investigator identified an uncertainty between two models: one, represented by the spike, in which the treatment had no effect (neither beneficial nor detrimental), and one in which the treatment had an effect (though not necessarily beneficial), of whose magnitude he was unsure.

A simple prior belief distribution can be constructed to capture such uncertainty between two models (M_0 and M_1): model M_0 allows no treatment effect so that $\beta = 0$ with probability 1; model M_1 allows a treatment effect with a non-informative (uniform) prior describing prior beliefs. Letting the prior beliefs in models M_0 and M_1 be p_0 and $1 - p_0$ respectively, then the prior odds for model M_0 versus model M_1 is $\lambda = p_0/(1 - p_0)$. Thus the pooled prior from the two models is a mixture distribution having a spike at the point of no effect with probability mass p_0 and is, for other effects, uniform with total probability mass $1 - p_0$. Note that Cornfield[8] suggested a similar prior distribution as a basis for hypothesis testing in clinical trials though he adopted a normal prior distribution for the treatment effect in model M_1.

The post-trial probability distribution $p(\beta|y)$ is then proportional to

$$p(\beta|y, M_0)\, p(M_0|y) + p(\beta|y, M_1)\, p(M_1|y). \tag{1}$$

Using the Bayes factor[9] B_{01}, defined as the ratio of posterior to prior odds for model M_0 versus model M_1, that is

$$B_{01} = \frac{p(M_0|y)/p(M_1|y)}{p_0/(1 - p_0)}, \tag{2}$$

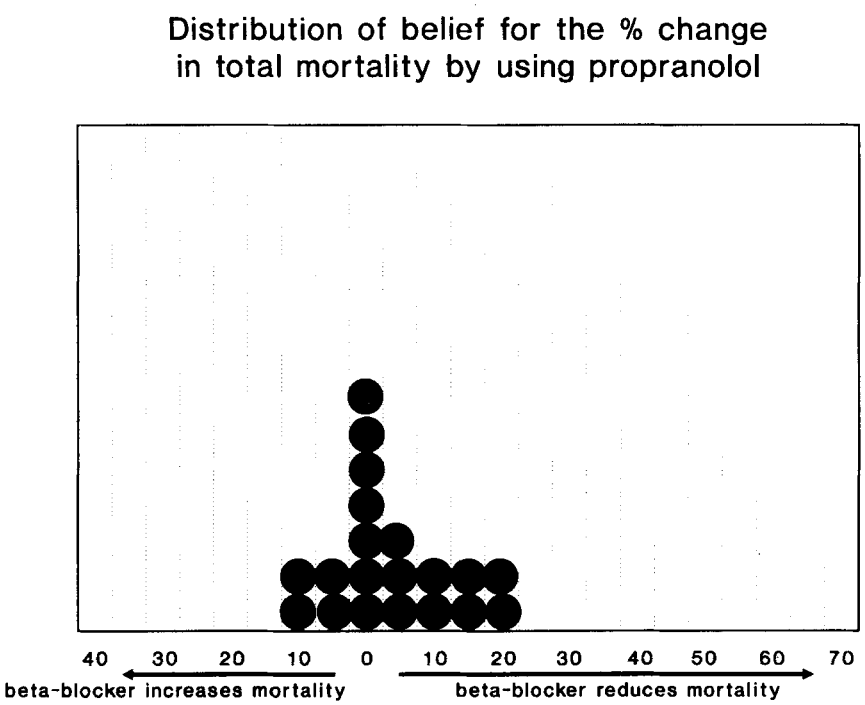

Figure 2. Pre-trial belief distribution of one investigator in the BAYREP trial showing the 'spike' of probability associated with no treatment effect

gives, with normalization,

$$p(\beta|y) = \frac{\lambda B_{01}}{1 + \lambda B_{01}}\, p(\beta|y, M_0) + \frac{1}{1 + \lambda B_{01}}\, p(\beta|y, M_1). \tag{3}$$

This is also a mixture distribution having a spike with probability mass $\lambda B_{01}/(1 + \lambda B_{01})$ at the point of no treatment effect (the left-hand term in the sum above) and elsewhere following the shape of the standardized likelihood distribution but with probability mass $1/(1 + \lambda B_{01})$ (the right-hand term). Thus characteristics of the sceptical clinician's post-trial beliefs can be related to their pre-trial belief p_0 that there is no treatment effect.

For analysis of the example, denote by l_P and l_T the observed log odds of dying in the placebo and treatment groups respectively. Then the Bayes factor for comparing models M_0 and M_1 is [10]

$$B_{01} = 2[\mathrm{var}(l_P) + \mathrm{var}(l_T)]^{1/2} \exp\{-1/2(l_P - l_T)^2/[\mathrm{var}(l_P) + \mathrm{var}(l_T)]\}.$$

When d_P of n_P patients randomized to placebo are observed to die, $\mathrm{var}(l_P)$ is estimated by $n_P/d_P(n_P - d_P)$, with a similar definition for $\mathrm{var}(l_T)$. For the AIMS trial, this gives a Bayes factor of 0·0011 independent of the choice of p_0. This factor describes how the trial data change the belief in model M_0 (no treatment effect) relative to that in model M_1 (treatment effect term included). For the AIMS trial, the data produce a substantial shift of belief away from model M_0 toward model M_1.

It is useful to plot the post-trial belief in a beneficial effect, $p(\beta < 0|y)$, versus the prior belief in no treatment effect, p_0. As β is zero with probability 1 in model M_0, equation (3) gives that

$$p(\beta < 0|y) = (1 + \lambda B_{01})^{-1}\, p(\beta < 0|y, M_1). \tag{4}$$

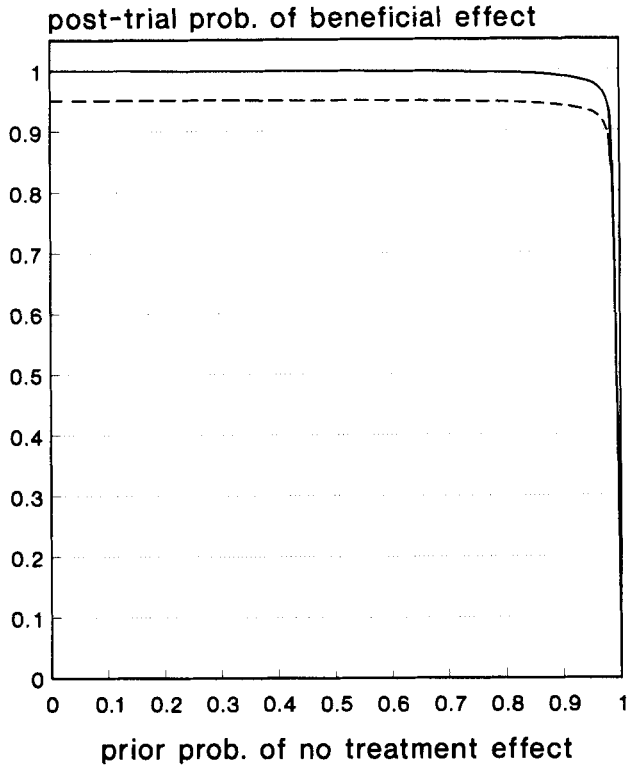

Figure 3. Relationship in the AIMS trial of the post-trial probability for a beneficial effect of anistreplase (bold line) and a reduction in the odds of dying by at least 20 per cent (dashed line) versus the pre-trial belief in no effect

The right-hand term is just the probability of a beneficial effect obtained from the standardized likelihood distribution adjusted by a factor to reflect the prior beliefs in no effect. Thus with no spike of prior belief in no effect, $\lambda = 0$, equation (4) gives the same post-trial belief of a beneficial effect as obtained from the standardized likelihood distribution. At the other extreme if, *a priori*, we have a spike of prior belief in no effect of probability one, then $\lambda = \infty$, and equation (4) gives zero post-trial belief in a beneficial effect, so that no amount of data can alter our pre-trial belief. Figure 3 shows the relationship between $p(\beta < 0|y)$ and p_0 for the AIMS trial (bold line). The proximity of the curve to a post-trial belief of one for a beneficial effect for all but the most extreme prior beliefs in no effect shows the strength of evidence coming from this trial. The data should overcome any realistic prior scepticism of the form described. The shape of the curve in Figure 3 follows from the magnitude of the Bayes factor: a value of one for B_{01} would give a linear relationship; values of B_{01} greater than one lead to a concave shape; values of B_{01} less than one lead to a convex shape.

Stronger evidence from a trial might be displayed by showing that substantial prior beliefs in no effect still lead to sizable post-trial beliefs in a larger beneficial effect. For example, also shown in Figure 3 (dashed line) is the post-trial belief in treatment producing a 20 per cent or greater reduction in the odds of dying. Akin to equation (4), this belief is given by

$$p(\beta < \ln(0{\cdot}8)|y) = (1 + \lambda B_{01})^{-1}\, p(\beta < \ln(0{\cdot}8)|y, M_1), \tag{5}$$

so explaining the similarity of the shapes of the two curves.

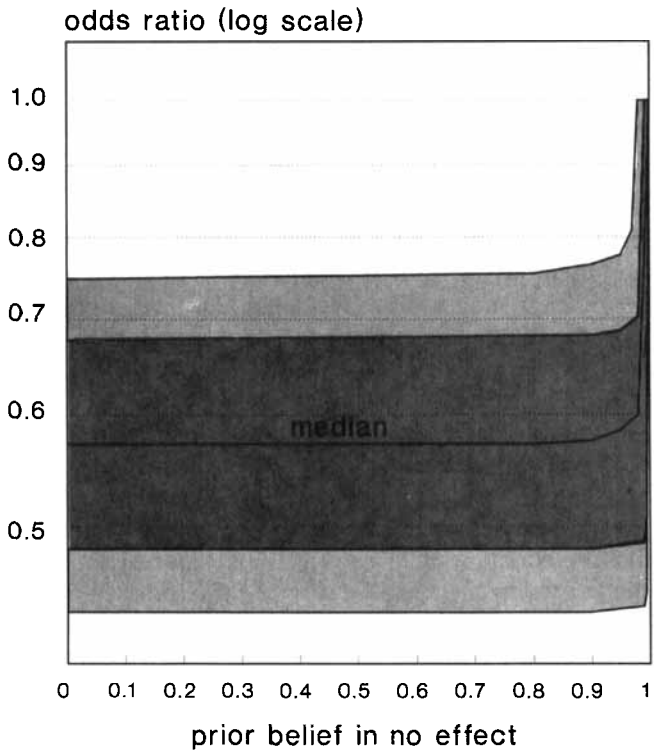

Figure 4. Relationship in the AIMS trial of the median effect and the 70 per cent (lighter shading) and 90 per cent (heavier shading) probability intervals for the effect of anistreplase versus the pre-trial belief in no effect

An alternative presentation of this analysis shows how probability intervals obtained from the probability distribution for β (given by equation (3)) depend on the prior belief in no effect, p_0. This enables the reader to focus on what he or she feels are clinically worthwhile treatment effects. Figure 4 shows these intervals for the AIMS trial: they are largely unaffected by the magnitude of the prior belief in no effect. The plot gives information about the shape and location of the standardized likelihood distribution as this is the post-trial probability distribution when $p_0 = 0$. Note that the intervals always converge to an odds ratio of one as p_0 tends to one.

5. EXAMPLE: HOW A REPORT MIGHT BE WRITTEN

In this Section I use the BAYREP trial to illustrate how Bayesian analyses might be reported in the medical literature. The report is restricted to the results for the survival endpoint, and is shorter than a complete trial report which would include details on study background, trial design and information concerning other outcomes. In the discussion section of this report, two other prior distributions are used to illustrate how the results of the trial might be interpreted in the context of the subjective opinions of the trial investigators and of previous results from earlier related studies.

Patients and methods

A total of 510 patients with oesophageal varices confirmed by biopsy were randomly assigned to receive either a beta-blocker ($n = 255$) or a matching placebo ($n = 255$). All patients had at least one varix of diameter 5 mm or greater, were over 16 years of age and received no beta-blocker agent during the month prior to randomization. Exclusion criteria were a history of upper gastrointestinal tract bleeding, regular use of beta-blocker or contraindications to their use.

All patients were followed for two years while receiving randomized treatment. The primary endpoint of the study was death. Comparison of groups is by intention-to-treat using a logistic regression model. Bayesian methods of analysis are used. A non-informative uniform distribution is used for the prior distribution for the log odds of dying for patients receiving placebo. This is assumed to be independent of the prior distribution used for the treatment effect of beta-blockers. In the Results section, a non-informative uniform distribution is used for the treatment effect. In interpreting the results (Discussion section), a mixture prior distribution is used to obtain Figure 6. This distribution is uniform apart from a probability spike of variable mass for a log odds ratio of 0. The distributions obtained in Figure 7 use prior distributions, one of which is based on a meta-analysis of similar trials reported in Hayes *et al.*[11] and assumes a common treatment effect in all trials; the other prior distribution reflects the pooled beliefs of this trial's investigators, derived using the approach of Spiegelhalter and Freedman[6] as adapted by Gore[7].

Results

Of the patients receiving beta-blockers, 40 (15·7 per cent) died compared with 55 (21·6 per cent) receiving placebo: an observed odds ratio of 0·68. Figure 5 shows the distribution of treatment effects that are supported by these data: 90 per cent (lighter shading) and 70 per cent (darker shading) probability intervals emphasize the effects having greatest support. The probability of an odds ratio less than 1, corresponding to a beneficial effect of beta-blockers, is 0·957; the probability of an odds ratio greater than 1, corresponding to an adverse effect on survival, is 0·043. The probability ratio of the treatment giving a benefit compared to an adverse effect is 22:1.

Discussion

The results of this study provide some support that primary prophylaxis with beta-blockers in cirrhotic patients with oesophageal varices reduces the odds of dying during two years of treatment, although the evidence is by no means conclusive. Whilst the results favour a beneficial effect on survival as compared with an adverse effect by a probability ratio of 22:1, more substantial effects on survival may be required in order to outweigh the established adverse effects of beta-blockers and so justify their use for primary prophylaxis. For example, when surveyed individually prior to this study, the largest odds ratios we felt acceptable before considering the use of beta-blockers in our routine practice ranged from 0·95 to 0·80. Support from this trial, shown by Figure 5, for such beneficial effects is weaker (a probability ratio of 14:1 for an odds ratio of 0·95 or better and a ratio of only 3:1 for an odds ratio of 0·80 or better).

Figure 6 shows how the median effect and the 70 and 90 per cent probability intervals are moderated for varying degrees of prior scepticism regarding whether beta-blockers

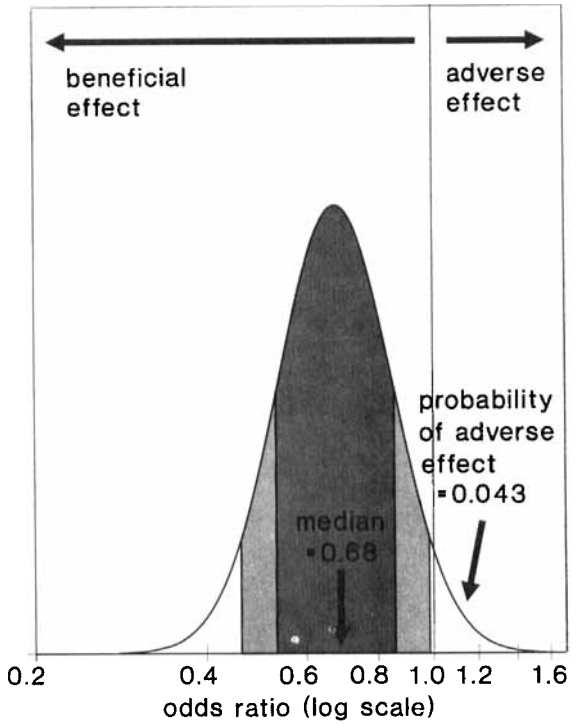

Figure 5. Distribution of effects of the beta-blocker relative to placebo having support from the data obtained in the BAYREP trial

have an effect on survival. Interested readers might quantify their own beliefs (held prior to reading this article) for beta-blockers having any effect on survival (adverse or beneficial) relative to their beliefs that they have no effect. For example, if prior belief in no effect is high, say 0·5, then the data do not move the median away from an odds ratio of 1 and the post-trial belief in a beneficial effect is weak (probability 0·31). Indeed, the upper bounds for the probability intervals shown include an odds ratio of one for all except those with very low prior beliefs (prior probability less than 0·06) in no treatment effect.

Figure 7 shows the results of updating both the pooled pre-trial beliefs of the study investigators and the distribution obtained from pooling the results of previously published studies.[11] As these distributions are very similar, the updated distributions are also similar. Compared with the standardized likelihood (also shown in Figure 7), these distributions discount the possibility of larger beneficial effects (for example, odds ratios smaller than about 0·5) but give slightly more support for the possibility of a beneficial effect (odds ratios of less than 1). However, there is still a reasonable probability that the beta-blocker effect is less than that considered worthwhile by the trial's investigators. For example, the post-trial belief for an odds ratio of greater than 0·9 ranges from 0·066 when using the prior distribution based on previous studies to 0·105 for the standardized likelihood.

In conclusion, although this study has substantially increased the experience from randomized controlled clinical trials regarding the use of beta-blockers for primary

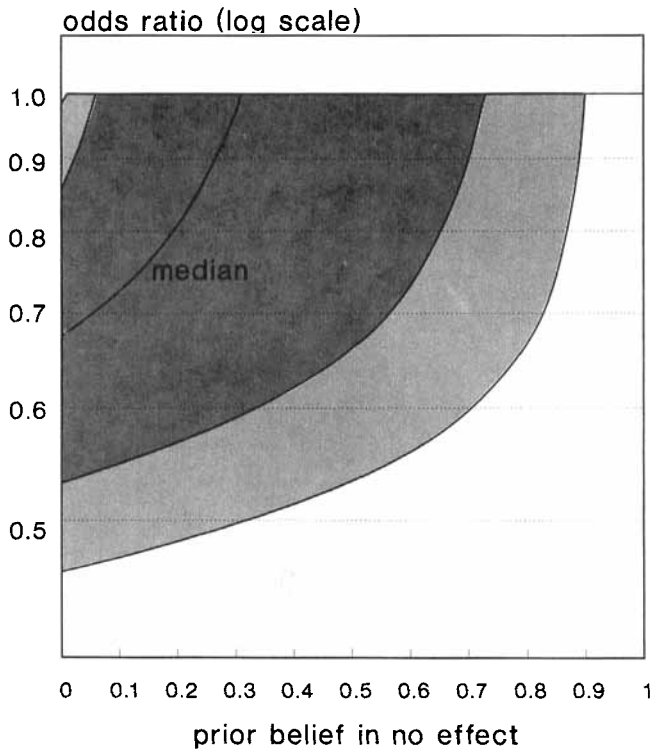

Figure 6. Relationship in the BAYREP trial of the median effect and the 70 per cent (lighter shading) and 90 per cent (heavier shading) probability intervals for the effect of the beta-blocker versus the pre-trial belief in no effect

prophylaxis, the evidence for a beneficial effect is encouraging but not yet sufficiently conclusive to recommend their routine use in clinical practice. Since our analyses do not discount the possibility of a moderate beneficial effect, further study is warranted.

6. DISCUSSION

Any clinical trial report should provide information in a form that can convince clinicians of a treatment's worth, or otherwise. Bayesian analyses can be valuable aids in achieving this: graphical displays of post-trial probability distributions allow clinicians to assess for themselves the likely size of effect and also illustrate sensitivity of interpretation to a range of prior beliefs. Comparison of corresponding figures for the AIMS and BAYREP trials (for example: Figures 1 and 5; Figures 4 and 6) show the differences between trials with strong versus weak evidence for a beneficial effect. However, if different choices of the prior distribution do affect interpretation, then it is important to avoid presenting only those which favour one particular view. Thus identifying appropriate prior distributions to illustrate sensitivity is crucial if Bayesian methods are to gain acceptance as aids in clinical trial reporting.

Using a non-informative, uniform prior distribution to give the standardized likelihood distribution stems from a desire that prior beliefs contribute very little to the inference relative to the data obtained. However, non-informativeness is not a straightforward concept and care must be taken before routinely adopting a uniform prior to express it. An important consideration is

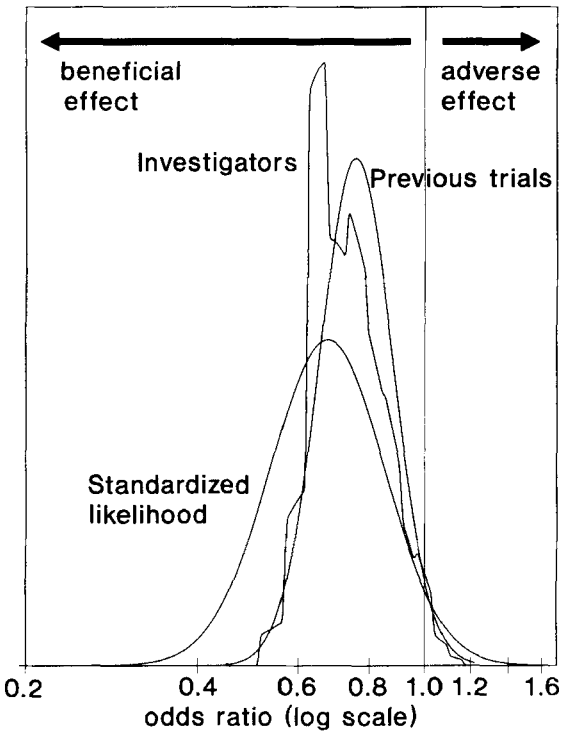

Figure 7. Post-trial probability distributions for the effect of the beta-blocker compared with placebo for prior distributions based on the investigators' pooled opinions and on the results from previous trials, showing also the standardized likelihood (as in Figure 5)

the scale on which such prior beliefs are expressed. In the AIMS and BAYREP examples, using a log odds ratio scale ensures that effects equally distant (on either side) from the point of no effect (a log odds ratio of zero) indicate equivalent differential effects of one treatment over the other, and also that the prior distribution is symmetric about that point. In contrast, the odds ratio scale does not have this symmetry about the point of no effect so a uniform prior distribution could not be considered non-informative. In general, a simple transformation will often be available to give a scale on which positive and negative departures of equal magnitude from the point of no effect imply an equal benefit of one treatment over the other, and on which prior probabilities can be distributed symmetrically about that point. Of course, there may be more than one scale which captures this concept of symmetry. In this case, expressing non-informativeness on the scale most naturally associated with practice would probably be preferable. Alternatively, the choice between scales might be motivated by one of the criteria put forward, for example, by Jeffreys[9] or Bernardo.[12] In a practical context, the choice will rarely make a marked difference to the probabilities obtained and the consequent inferences drawn, unless the study is particularly small. In the latter case, the interpretation of the study will also be very sensitive to different degrees of scepticism.

When using non-informative improper prior distributions, the integral of the likelihood function may not be finite and so the standardized likelihood distribution cannot be obtained. In most practical applications this is unlikely to be a problem. If it does arise, then presentation of the likelihood function itself would guide the reader as to the range of values that are supported

by the data. Proper prior distributions can then be used to illustrate the sensitivity of the conclusions to different forms of prior belief.

The use of a mixture prior distribution which is 'spiked' at the point of no treatment effect allows analyses that illustrate sensitivity to different levels of prior scepticism. This gives graphs that are easily interpreted showing the strength of the results to overcome prior scepticism. These also allow readers to assess objectively the impact of the results on their own beliefs. Similarly, the investigators can use these graphs as an aid in discussing the results. Such discussion requires consideration of what degree of prior scepticism might exist in practice. Berger and Delampady[13] consider in detail the use of the spiked prior in testing point hypotheses and attach a prior probability of 0·5 to the spike. This leads them to conclude that the strength of evidence required to move opinion away from the point hypothesis of no treatment effect needs to be substantially larger than is suggested by classical p-values. In the discussion of that paper, Casella and Berger express an alternative view that, in undertaking an experiment, researchers expect *a priori* to find a difference and so a value closer to zero (they suggest 0·1) is more appropriate. However, in reporting clinical trials the views of practitioners not involved in the trial are important. Even with phase III trials, where there may be encouraging evidence from earlier phase trials, there may be substantial prior scepticism of a treatment's effectiveness amongst clinicians external to the trial. This may be particularly true amongst clinicians treating diseases, such as cancer, where rather few of the many treatments that are tested are found to be effective in clinical practice. Thus it is important to show the full range of prior probabilities from zero to one that are associated with the spike, and discussion based on moderate prior probabilities (say, 0·3 to 0·5) would be appropriate except in diseases where therapeutic progress is particularly slow (when higher levels or prior scepticism would be important) or where it is more rapid (when lower degrees might be sufficient).

A related issue concerns the choice of prior distribution used to describe beliefs under model M_1, in which a treatment is believed to have some effect. In this paper, a non-informative uniform prior has been used as this gives a link to the analysis based on the standardized likelihood distribution. However, individuals having prior beliefs under model M_1 that favour some sizes of effect more than others may be more readily convinced of a treatment's effect than individuals who describe their prior beliefs under model M_1 using the non-informative prior. This is true even if that individual's prior is symmetric about the point of no effect because smaller observed effects are consistent with his beliefs under model M_1 and so enable an easier transition from model M_0 to model M_1. On the other hand, individuals having a prior belief under model M_1 that is predominantly in the opposite direction to that suggested by the data would have a post-trial belief in a beneficial effect that is less strong than portrayed in figures such as 3, 4 and 6. Therefore the use of the non-informative prior is an intermediate choice. It provides a simple approach for assessing the strength of the results to overcome prior scepticism and requires that only one parameter be varied in order to describe the level of scepticism.

Other, possibly more controversial, prior distributions might be considered. For instance, in the BAYREP example, prior distributions derived from combining results from similar previous trials and also from pooling the investigators' own subjective opinions were used. However, the derivation of these prior distributions may raise many questions. Using previous trial results is effectively conducting a meta-analysis, and may raise questions about which trials were included and the type of model used (fixed or random effects). This might detract from the emphasis on reporting the current trial's results. Suspicions that investigators may be overly optimistic about a treatment's effect may lead to controversy if subjective opinions are used. Further problems may arise when clinicians involved in a trial fully understand the impact that their 'prior opinions' can have on this particular presentation of the results. Thus, whilst illustration of the sensitivity of

the conclusions drawn from using such priors may be helpful, this should be restricted to the discussion section of the paper.

In conclusion, the methods described enable a clinical reader to more fully assess the support from the data of a trial for a treatment's effectiveness than is provided by classical methods of analysis. A useful first step in moving toward acceptance of Bayesian analyses in trial reports would be the presentation of trial results using standardized likelihood distributions. This is unlikely to raise undue controversy amongst statisticians because of the use of a non-informative prior and would be intuitively appealing to clinicians because of the simplicity of understanding portrayed by probability intervals and tail probabilities. However, it would represent a major shift in presentation requiring careful guidance and education for clinical readers. In this respect an important objective of this paper is to generate discussion amongst statisticians and clinicians. It is important to realize the advantages of using the Bayesian approach while avoiding the problems that will arise if analyses are presented with a variety of ill-defined prior distributions.

ACKNOWLEDGEMENTS

I am grateful for the interest taken by clinicians involved in the beta-blocker trial in completing questionnaires about their beliefs. I am also grateful to Simon Thompson, Stuart Pocock, and Els Goetghebeur and the referees for useful comments on earlier versions of this paper. This work was supported in part by grant number CA39640 from the National Cancer Institute.

REFERENCES

1. Spiegelhalter, D. J. and Freedman, L. S. 'Bayesian approaches to clinical trials', in Bernardo, J. M., DeGroot, M. H., Lindley, D. V. and Smith, A. F. M. (eds.), *Bayesian Statistics 3*, North-Holland, Amsterdam, 1988.
2. Racine, A., Grieve, A. P., Fluhler, H. and Smith, A. F. M. 'Bayesian methods in practice: experiences in the pharmaceutical industry', *Applied Statistics*, **35**, 93–150 (1986).
3. AIMS Trial Study Group. 'Long-term effects of intravenous anistreplase in acute myocardial infarction: final report of the AIMS study', *Lancet*, **335**, 427–431 (1990).
4. Bosch, J., Groszmann, R. J., Grace, N., Navasa, M., Mastai, R., Conn, H. O. and Rodes, J. 'Propranolol in the presentation of first haemorrhage from oesophageal varices: results of a randomized, double-blind, cooperative clinical trial', *Journal of Hepatology*, **7** (supplement 1), S12 (1988).
5. Pocock, S. J. and Hughes, M. D. 'Estimation issues in clinical trials and overviews', *Statistics in Medicine*, **9**, 657–667 (1990).
6. Spiegelhalter, D. J. and Freedman, L. S. 'A predictive approach to selecting the size of a clinical trial, based on subjective opinion', *Statistics in Medicine*, **5**, 1–13 (1986).
7. Gore, S. 'Biostatistics and the Medical Research Council', *Medical Research Council News*, 19–20, June 1987.
8. Cornfield, J. 'A Bayesian test of some classical hypotheses – with applications to sequential clinical trials', *Journal of the American Statistical Association*, **61**, 577–594 (1966).
9. Jeffreys, H. *Theory of Probability*, Oxford University Press, Oxford, 1961.
10. Spiegelhalter, D. J. and Smith, A. F. M. 'Bayes factors for linear and log-linear models with vague prior information', *Journal of the Royal Statistical Society, Series B*, **44**, 377–387 (1982).
11. Hayes, P. C., Davis, J. M., Lewis, J. A. and Bouchier, I. A. D. 'Meta-analysis of value of propranolol in prevention of variceal haemorrhage', *Lancet*, **336**, 153–156 (1990).
12. Bernardo, J. M. 'Reference posterior distributions for Bayesian inference', *Journal of the Royal Statistical Society, Series B*, **41**, 113–147 (1979).
13. Berger, J. O. and Delampady, M. 'Testing precise hypotheses', *Statistical Science*, **2**, 317–352 (1987).