the start of a trial, even if they are only to be used informally. Contrasting prior beliefs with accumulating data can provide a means of identifying over-optimistic expectations, but this can only take place if those expectations have been explicitly recorded. Sometimes prior expectations can be dead on: the Beta-blocker heart attack trial (BHAT) was designed around an expected 28% drop in mortality derived from previous studies (BHAT Research Group, 1984); after 3837 patients had been randomized the observed improvement was exactly 28%! (See BHAT Research Group, 1987.) It would be rather optimistic to think that all prior judgments will be so accurate, especially when similar studies have not been carried out and one is reliant on purely subjective opinion.

An encouraging sign is the willingness of established researchers in clinical trials to take the Bayesian approach seriously: Armitage (1988, 1989) illustrates many of the points made in this discussion, while Pocock and Hughes (1990) provide details of Bayesian estimation following early termination of a trial in order to overcome the excessive bias of the standard estimate. We feel confident that slow but steady progress toward Bayesian design and monitoring will continue to be made in the future, and feel sure that Professor Breslow's paper will help in this regard.

Comment

M. Zelen

Bayesian methods have influenced our thinking about the foundations of statistical inference but have not enjoyed widespread popularity in applications. Professor Breslow's paper is a fine summary of some of the settings in which Bayesian methods have been applied with success to real data problems. The paper serves as a reminder that Bayesian methods are beginning to be utilized in the analysis of data arising in the health sciences. I would expect this trend to increase as Bayesian software becomes more available. However, even with access to appropriate software, the increased use of Bayesian methods will be dampened by the sensitivity of these methods to model specificity. A widely prevailing view is that inferences should rely on reasonably robust procedures. As a result, one is likely to see Bayesian methods applied to situations which have insufficient data to make frequency-based inferences or to situations which directly arise from Bayesian considerations. It is this latter remark on which I will comment further.

The Bayesian philosophy seems particularly appealing and appropriate in case–control settings. This methodology is aimed at inferring whether exposure to a potential causal factor is associated with the incidence of a particular disease. Starting from a collection of cases and controls, one must infer if the case exposure to a causal factor is "unusual" when compared to controls. One can use the information from a control group to calculate the posterior distribution of exposure. In many instances, there may be so much prior information available about exposure of a population (e.g., lifestyle habits of smoking and drinking, etc.) that the limited information available from a sample of controls may generate a posterior distribution of exposure which is nearly the same as the prior distribution. In such situations, one can carry out an analysis of the cases and their exposure without even generating data on a control group. The frequentist view of case–control studies does not permit studies without controls. This represents a serious shortcoming of the frequentist methodology for case–control studies. To cite an extreme example, suppose one has a potential causal factor which is rare in the population, yet the available cases all have been exposed to the causal factor. It would be ludicrous to carry out a case–control study. Yet this is what the frequency point-of-view dictates.

The frequentist model for case–control studies is that random samples are drawn from a population of cases and controls. In practice, this assumption is unrealistic and is rarely met in practice. Data on cases are usually drawn from hospitals, registries or whatever data collection mechanism would yield a convenient set of cases. Controls may be gathered in a variety of ways, but often it is not at all clear if the controls are from the same population as the cases. Various matching techniques are used to attempt to make cases and controls comparable, but there is no way to account for unknown factors which can influence

M. Zelen is Professor of Statistics and Chairman of the Department of Biostatistics at Harvard University. His mailing address is Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115.

exposure and disease incidence. Controversies abound in the applications of the case–control methodology due to questions on the comparability of cases and controls. These studies are easy to fault.

The logic of Bayesian methods for case–control studies is overwhelming. Aspects of this methodology and point-of-view are discussed in Zelen and Parker (1986).

Rejoinder

Norman Breslow

My choice of the topic "Biostatistics and Bayes" for the ASA session on historical perspectives and new directions in biometry was motivated by several concerns: to attempt to avoid the tedium that often accompanies "past, present and future" overviews; to use the opportunity to undertake some remedial self education in an area that I had long neglected; and to simulate both myself and my colleagues to glance up from the project work that often so completely absorbs us to consider some broader issues. Having anticipated criticism from both biostatisticians and Bayesians for daring to comment on a controversial topic to which I had personally contributed no research work, I was pleased with the generally supportive remarks of the discussants and perhaps even a little disappointed that I was not called more severely to task. I appreciate their efforts in contributing to the discussion and I thank the editors for their indulgence and encouragement.

PROBLEMS WITH MANY PARAMETERS

Professor Armitage and Dr. Jennison, while generally sympathetic to the use of (empirical) Bayes methods for problems involving large numbers of unknown but related parameters, both remark that classical random effects models are available outside the Bayesian framework. Reduction of the estimation problem to one involving a small number of parameters that describe means and variance components indeed is a necessary first step. However, in the standard set-up, the "random effects" often are considered more of a nuisance that complicate inferences on the main effects. The appealing feature of the Bayes approach is its focus on joint shrinkage estimation of the original parameters, i.e., the random effects, for presentation in maps or reconstructed images. Use of the Bayesian paradigm as a technical tool for sharpening estimates of a large number of parameters considered in the aggregate seems well established and noncontroversial. Viewed from this perspective, Bayesian methods may have an even greater role to play in exploratory data analysis than Dr. Flühler would acknowledge. It is only when one needs to single out one of the estimated parameters for special inference and decision that greater caution in assessing model assumptions, including specification of the prior, is called for. This underlies my own hesitation in fully embracing the use of Bayes techniques for cancer risk assessment and, evidently, Professor Armitage's hesitation in applying them to bioassays of pharmaceutical compounds.

BIOEQUIVALENCE

I am grateful to Drs. Spiegelhalter and Freedman for helping, both in their commentary and in several published articles, to clarify my thinking about bioequivalence, especially in the context of clinical trials. They correctly emphasize the need to keep the specification of the indifference region, or range of equivalence, quite separate from consideration of prior beliefs about the actual difference between treatments. Westlake, Armitage and others have shown how the use of confidence intervals superimposed on the line graph of preference and indifference regions provides a reasonable frequentist solution to the problem of inferring equivalence in simple problems. In reply to Dr. Jennison, I did not mean to imply that adequate frequentist solutions could not be found more generally. I agree with Drs. Spiegelhalter and Freedman, however, that the Bayesian interpretation in terms of the posterior probability of equivalence is more natural and, knowing how they are inclined to think of confidence intervals, feel nearly certain that my medical associates would agree also.

To amplify on Dr. Flühler's remarks about the role of Bayesian thinking at the design stage, I would again like to borrow from Spiegelhalter and Freedman's commentary and suggest to Dr. Jennison that his OC curve be averaged with respect to a plausible prior distribution for the treatment differences in order to decide if the proposed study has any hope of achieving its goal. The prior used for this calculation, which would indicate the pharmaceutical company's belief in the efficacy of its new drug, might well differ