ROSENBAUM, P. R. and RUBIN, D. B. (1984). Sensitivity of Bayes inference with data-dependent stopping rules. *Amer. Statist.* **38** 106–109.

SELWYN, M. R., DEMPSTER, A. P. and HALL, N. R. (1981). A Bayesian approach to bioequivalence for the 2 × 2 changeover design. *Biometrics* **37** 11–21.

SELWYN, M. R. and HALL, N. R. (1984). On Bayesian models for bioequivalence. *Biometrics* **40** 1103–1108.

SIEMIATYCKI, J., GERIN, M., RICHARDSON, L., HUBERT, J. and KEMPER, H. (1982). Preliminary report of an exposure-based, case-control monitoring system for discovering occupational carcinogens. *Teratogenesis Carcinog. Mutagen* **2** 169–177.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

SIMON, R. (1977). Adaptive treatment assignment methods and clinical trials. *Biometrics* **33** 743–749.

SIMON, R. (1982). Patient subsets and variations in therapeutic efficacy. *Br. J. Clin. Pharmacol.* **14** 473–482.

SMITH, A. F. M., SKENE, A. M., SHAW, J. E. H. and NAYLOR, J. C. (1987). Progress with numerical and graphical methods for practical Bayesian statistics. *The Statistician* **36** 75–82.

SMITH, A. F. M., SKENE, A. M., SHAW, J. E. H., NAYLOR, J. C. and DRANSFIELD, M. (1985). The implementation of the Bayesian paradigm. *Comm. Statist. A—Theory Methods* **14** 1079–1102.

SPIEGELHALTER, D. J., FREEDMAN, L. S. and BLACKBURN, P. R. (1986). Monitoring clinical trials: Conditional or predictive power? *Controlled Clin. Trials* **7** 8–17.

TAKASUGI, M., TERASAKI, P. I., HENDERSON, B., MICKEY, M. R., MENCK, H. and THOMPSON, R. W. (1973). HL-A antigens in solid tumors. *Cancer Res.* **33** 648–650.

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82** 528–540.

THOMAS, D. C., GOLDBERG, M., DEWAR, R. and SIEMIATYCKI, J. (1986). Statistical methods for relating several exposure factors to several diseases in case-heterogeneity studies. *Statist. in Medicine* **5** 49–60.

THOMAS, D. C., SIEMIATYCKI, J., DEWAR, R., ROBINS, J., GOLDBERG, M. and ARMSTRONG, B. G. (1985). The problem of multiple inference in studies designed to generate hypotheses. *Amer. J. Epidemiol.* **122** 1080–1095.

TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86.

TSIATIS, A. A., ROSNER, G. L. and MEHTA, C. R. (1984). Exact confidence intervals following a group sequential test. *Biometrics* **40** 797–803.

TSUTAKAWA, R. K. (1988). Mixed model for analyzing geographic variability in mortality rates. *J. Amer. Statist. Assoc.* **83** 37–42.

TUKEY, J. W. (1977). Some thoughts on clinical trials, especially problems of multiplicity. *Science* **198** 679–684.

WESTLAKE, W. J. (1979). Statistical aspects of comparative bioavailability trials. *Biometrics* **35** 273–280.

WHITEHEAD, J. (1983). *The Design and Analysis of Sequential Clinical Trials*. Ellis Horwood, Chichester.

ZEGER, S. L. and LIANG, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42** 1–10.

Comment

Peter Armitage

This persuasive paper should be welcomed by all biostatisticians, not least because the author succeeds in conveying his enthusiasm for (although to some extent his reservations about) Bayesian analysis without indulging in the Messianic fervor so characteristic of some of its proponents. For my part I am convinced that Bayesian methods have a major role to play in the analysis of biomedical data, although I am as skeptical about claims that they provide an all-embracing "world-view" of statistics as I am about similar claims in the realms of politics, art or religion.

Since Dr. Breslow starts with some fascinating autobiographical detail, it may not be out of place to add a few personal comments. When I entered medical statistics in 1947, the discipline was still struggling to take on board the pre-war advances of Fisher and his contemporaries. In Britain, J. O. Irwin was, among biostatisticians, almost a lone representative of the Fisher–Neyman traditions; in the United States, Cochran had yet to enter biostatistics and take on his important leading role. (The developments of the 1920s and 1930s were, of course, more deeply established in agricultural research.) Jeffreys appeared as a lone figure of great stature but almost completely lacking in influence. Bayesian methods were propounded, in the U.K., by a few people, including W. Perks, an actuary, and I. G. Good, but to little effect, and it was not until the appearance of L. J. Savage's book in 1954 that more than a handful of statisticians took Bayes seriously.

In the gradual process of consolidating the use of "standard" methods, most of us gave little thought to the apparently more formidable task of introducing Bayesian inference and decision theory. I must have been one of the English statisticians, during Norman

Peter Armitage is Emeritus Professor of Applied Statistics at the University of Oxford. His mailing address is 71 High Street, Drayton, Abingdon, Oxon OX14 4JW, England.

Breslow's very welcome visit, who, he says, openly exhibited "skepticism about decision theory." I am not sure that I am much less skeptical now, apart from consideration of a few well-defined situations. But Bayesian inference without the problem of defining utilities is a different matter and it is primarily this topic that forms the basis of Breslow's paper.

I think we can distinguish a number of distinct scenarios for which Bayesian methods might be contemplated. First there is the sort of situation exemplified by Sections 3 and 4 of the paper, where many different entities (areas, treatments, sites, etc.) are characterized by an exchangeable set of parameters. The popular approach is empirical Bayes (which the complete Bayesian would regard as an approximation to true Bayes), leading typically to shrinkage of estimates. Many problems of this sort can be expressed in a non-Bayesian framework in terms of components of variance, and an awareness of "regression to the mean" (for instance in dealing with extreme blood pressure measurements) has a long history. But the empirical Bayes framework is convenient and penetrating. I, and I think most biostatisticians, find this approach persuasive, and I believe that it will be widely used in the future. (A promising area not mentioned by Dr. Breslow is that of overviews or "meta-analyses" of clinical trials.) There are, of course, points to debate—for instance in the extent to which different entities should be regarded as exchangeable, and in the form of the prior distribution. A nonparametric prior seems to avoid unnecessary assumptions but gives the highly implausible point distributions as ML estimates.

Why has this approach not been used much until the last decade or so? I imagine that in the first few pre-war decades medical research provided many fewer instances of large-scale studies in which these sets of mutually relevant entities appeared. Early examples would have been mortality rates (where the sort of methods described by Clayton and Kaldor could have been used in the study of regional or occupational differences) and bioassays of pharmaceutical preparations. But in this latter example, it would be quite tricky to argue that a preparation, which on standard analyses fails a specification, should be released to the public on the grounds that most other preparations have passed the test.

A second type of scenario is that in which clear decisions are to be taken after analysis of data. An example would be in-house testing of a drug during the course of its development by a pharmaceutical company. Even though utilities may be formulated only roughly, it seems entirely reasonable that the company should introduce its own subjective priors, in taking a decision that affects its own interests. Another situation is that of a clinical trial carrying clear decision implications for the investigators; again, it is reasonable that they should grope toward a definition of their own priors and utilities (Spiegelhalter and Freedman, 1988), even if they might differ from everyone else's.

There are other scenarios where I am less persuaded. These are, broadly, scientific studies to elicit information on some question where there is no convincing exchangeable set from which to draw strength, where no clear decisions will follow and the consequences of various outcomes are too diffuse to formulate simply, and where the primary purpose of analysis is to present the implications as simply and as objectively as possible. I think that most laboratory experimentation falls in this category, as do many clinical trials and epidemiological surveys. Now these could certainly be put through the Bayes mangle, perhaps with duly flattened priors, or perhaps by exhibiting only the likelihood so that users could use their own priors. The information conveyed would usually be similar to that conveyed by standard non-Bayesian means, and statisticians who believe they they can communicate better with the scientific public by standard methods should be encouraged to do so. This I take to underlie the conciliatory remarks in Dr. Breslow's last sentence.

There remains the thorny question of interim analysis in clinical trials. The Bayesian sees no problem here, except that of converting the stubborn clinical trial statistician who is concerned about the frequency effects of repeated testing of the data. In my experience medical investigators find this a natural concern, regarding it in much the same light as other examples of multiplicity like those considered by Dr. Breslow. I have in the past spent a good deal of effort examining the frequency properties of various types of stopping rule. I an now inclined to think that too much fine detail in this area of research is ill-advised, because in most of the large-scale trials now undertaken, for example for the treatment of cardiovascular disease, a decision whether or not to stop at some stage depends on many factors, only one of which is the question of boundary crossing. I would be content to quote nominal P-values, or some near-equivalent, with a description of the stopping procedure and some cautionary remarks about the broad interpretation of the tests. From this point of view, the existing work on frequency properties of stopping rules would be regarded as providing a guideline to the sort of effect that might be expected, rather than a way of getting precisely adjusted P-values.

Incidentally, I am not persuaded by the arguent at the end of Section 6, that the problem of optional stopping disappears if one realizes that the

appropriate prior should have a probability mass at the null value. It would be quite reasonable to have a continuous prior distribution function, but nevertheless to be interested in absorption probabilities at the null, as giving a bound to those at non-null values.

I have ignored in my comments many parts of Dr. Breslow's paper which are of great interest, for instance his remarks about model selection with particular reference to risk assessment. The paper will continue to stimulate interest for many years to come.

Comment

H. Flühler

Breslow's paper is a most interesting account of the Bayesian approach to solving problems in a biological context. Although his exposition does not rely and is not supported by practical experiences of his own, he presents manifold biological and medical application problems which were preferably tackled by applied statisticians from a Bayesian point of view. I support the message of Breslow's survey that progress in the statistical sciences is achieved most efficiently by a mature integration of the Bayesian thinking in applications.

Of the many topics deserving discussion I shall concentrate on three. First, I shall make some general comments about the Bayesian impact—from the perspective of an applied statistics unit in a major chemical and pharmaceutical company—to the various stages of statistical activities. Second, I shall address the topic of longitudinal data analysis, because I feel that the Bayesian approach will offer a most dramatic progress to all types of hierarchical models—supposing the workable tools which are underway will become available to the practitioner. Third, I shall refer to the interpretational and predictional flexibility offered by the Bayesian paradigm to the scientists in making inferential assessments based on experimented evidence.

GENERAL COMMENTS

The application of statistics is the basic foodstuff for progress. In order to achieve good statistical thinking and analysis, the scientific context has to be considered and understood. The multidisciplinary collaboration stimulates novel and unconventional approaches in solving statistical problems. Four different stages in the scientific learning cycle are identified, namely (i) the informal and less structured framework; (ii) the design phase; (iii) the reporting of inferences; and (iv) the diagnostics and model criticism.

Breslow's paper mainly outlines the impact and benefits of the Bayesian approach in stages (iii) and (iv) which I fully agree. However, the practical statistician is exposed to all four stages in any sequence and repetitive cycle.

Exploratory data analysis methods combined with interactive high density dynamic graphics and classical dimension reduction techniques are the essential ingredients for the practical statistical activity of stage (i). A recent account is presented by Weihs and Schmidli (1990) in this journal. Intuition and a free mental framework in respect to modeling and searching for structure are the characteristic elements of this activity phase. The Bayesian thinking, however, which requires a more or less structured framework, does play a minor role in this context. Prior information in respect of the application background and statistical expertise are essential components applied by the practitioner in an informal way.

How does the Bayesian framework support stage (ii), the design phase? Prior knowledge should always be available at the design phase assuming the scientific investigation as an on-going learning process which involves an iterative cycle of design, experiment, analysis and interpretation. The available prior information is applied and imbedded into the design phase in a more informal, natural thinking process. A formal procedure is presented by Hedayat, Jacroux and Majunder (1988) for comparing treatments with controls.

Bayesian methodology however strongly supports the reporting inferences process, stage (iii), and the diagnostics and model criticism, stage (iv). A theoretical account of the potential power is given by Smith (1986). The Bayesian paradigm could, however, play a much stronger role in a practical context. Why do these methods not get off the ground? First, there is an obvious educational deficiency in Bayesian methods. Second, many statisticians apply a philosophical

H. Flühler is the Head of the Mathematical Applications section of the CIBA-GEIGY Company, CH-4002 Basel, Switzerland.