

Comment

Susan S. Ellenberg

Dr. Fleming's discussion of some issues currently facing statisticians who are actively involved in clinical research is both interesting and timely. His paper provides excellent examples of important methodological research stemming from day-to-day collaboration between clinicians and statisticians.

The discussion of data monitoring committees is most welcome, as such committees are becoming an increasingly standard feature of large multicenter trials, and yet there has been relatively little in either the statistical or medical literature about their operational aspects. Attempts to standardize the monitoring of clinical trials along the lines described by Dr. Fleming have been particularly controversial in AIDS trials (Ellenberg, Myers and Hoth, 1993). In particular, some pharmaceutical companies have objected to their lack of access to interim comparative data, on the grounds that this policy will lead to delays in making effective drugs available and in initiating new research projects that would be based on results of preceding studies. Although this is certainly a legitimate concern, the potential conflict of interest may be of even greater concern. It could well happen, for example, that a company provided regularly with interim results of a government-sponsored trial could decide at some point that the data were sufficiently strong to support marketing approval for its drug; an independent board, on the other hand, might believe that the data were insufficiently definitive for purposes of Public Health Service recommendations. If the company proceeded to file for marketing approval with the Food and Drug Administration based on the interim results, these results would have to be made public, which would make it difficult or impossible to complete the trial as planned.

The atmosphere in cancer trials is quite different. The example of Intergroup 0035, the study of 5-FU + levamisole as adjuvant therapy for colon cancer, is a case in point. This study was mounted specifically to confirm the results of a smaller study. The latter results were widely known in the oncologic community, had been presented at major scientific meetings and were published during the course of the intergroup trial. There was no question concerning the quality of earlier

study. Yet the Intergroup trial had no trouble entering a large number of patients, one-third of whom were randomized to receive no adjuvant treatment. Such a situation is unimaginable in AIDS, where reporting networks within the patient community are so extensive that promising results from early phase I studies have led to widespread adoption of new treatment regimens even prior to the publication of these early studies. The intense interest of both the patient and the scientific communities in the potential of new agents to treat AIDS and AIDS-related disorders is due in large part to the small number and limited efficacy of the currently available agents. This extraordinary sense of urgency enhances the need to maintain strict confidentiality of interim results if there is to be any hope of completing studies that will provide reliable direction to physicians and patients.

Fleming's comparison of the early stopping of ACTG 019 with that of the Intergroup cancer study is important for understanding yet another major difficulty in studying treatments for AIDS. Unlike the adjuvant therapy administered for a limited time period in the Intergroup study, many AIDS drugs (antiretroviral drugs as well as prophylaxis for opportunistic infections) are chronic therapies, taken (in principle) for the remainder of one's life. Because no current therapies are curative, perhaps slowing but not reversing or even stabilizing the course of disease, and nearly all have undesirable side effects, patients are easily motivated to abandon an ongoing medication when a new therapy appears on the scene with some hope of being shown superior to current drugs, even when this hope is based on very limited data. Given the nature of the disease, it is difficult to be unsympathetic with these patients; nevertheless, it is clear that assessing the effects of these treatments on survival (the ultimate endpoint of interest for a fatal disease) will be highly problematic if patients remain on their assigned treatment for only a relatively short time—say, 6 months of a remaining lifetime of 5–7 years. The difficulty of maintaining patients on assigned study medication has led many clinical investigators to believe that we will never be able to perform comparative evaluations of survival in patients treated at early stages of disease with any acceptable degree of reliability.

The problems associated with assessing survival have led investigators to search for earlier outcomes on which to base evaluations of new therapeutic approaches. Fleming has described in convincing detail the issues surrounding the use of such "surrogate markers"

Susan S. Ellenberg is Chief, Biostatistics Research Branch, Division of AIDS, National Institute of Allergy and Infectious Diseases, 6003 Executive Blvd., Rm. 2B27, Bethesda, Maryland 20892.

as primary clinical trial endpoints. Although I fundamentally share his concerns, I would cite one important aspect of the AIDS context that is relevant to the use of such endpoints in trials designed to obtain marketing approval for new drugs. HIV-infected patients, unlike patients recovering from myocardial infarction or suffering from chronic granulomatous disease, will inevitably die of their disease within a short time relative to their otherwise expected remaining lifetime. The best we can hope for from current therapies is a modest to moderate prolongation of survival. In this circumstance, it does not seem inappropriate to accept a higher level of risk in deciding what therapies might be made available. Whether therapies that have only shown positive effects on early markers should be distributed in "expanded access" or "parallel track" programs, or whether the FDA should permit their manufacturers to market them, may be more of an economic than a scientific issue. Whatever mechanism is used, it will ultimately fall to federally funded research programs of the Public Health Service to mount trials that compare available regimens and move toward defining optimal treatment strategies for patients at various stages of disease. In these trials, it will be essential to study clinical efficacy—that is, physical rather than laboratory manifestations of disease—until and unless we discover markers that come much closer to meeting the Prentice criteria.

It is encouraging to learn of the innovative investigations by Fleming and colleagues of the potential use of the auxiliary information present in early markers of disease to strengthen evaluation of therapies when only limited long-term clinical data are available. As Fleming notes, the circumstances under which this type of approach will significantly add to our ability to assess treatments reliably are somewhat limited. Nevertheless, it would be of interest to test out such approaches in data sets in which the relationship between the surrogate and the "true" endpoint is fairly well characterized—for example, if S were blood pressure and T were heart attack or stroke. The problem is complicated in AIDS because there has been experience with relatively few treatments and therefore little data regarding the correlation between S and T in the presence of different therapies. If this correlation varies greatly according to the particular regimen being administered, it would be difficult to use this approach in any routine way.

In conclusion, I would like to congratulate Dr. Fleming for highlighting some of the issues biomedical statisticians are struggling with, and hope that his paper will inspire more statisticians to become actively involved in, and even leaders of, the process of planning and carrying out medical research programs.

Comment

Vern T. Farewell and Richard J. Cook

INTRODUCTION

In this paper, Dr. Fleming provides an excellent review of some current methodological problems facing health scientists involved in clinical trials. Some issues considered in detail are monitoring clinical trials, the analysis of equivalence trials, multiple endpoints and surrogate markers. We will remark on each of these in turn.

MONITORING

The examples cited clearly demonstrate the importance of a monitoring committee for moderate to

large-scale sequential clinical trials. In particular, a specialized and centralized Data Monitoring Committee (DMC) for the AIDS Clinical Trial Group (ACTG) is discussed. Such a specialized monitoring committee has immediately obvious advantages. As more trials are passed through the DMC, the disease-specific knowledge gained from early trials can be applied to later studies.

In principle, there are a variety of other diseases that require DMCs. For fields with less trial activity and experience, it may be advantageous to provide access to less specialized DMCs. Although it may be necessary to supplement the available expertise for individual trials, this more general DMC could provide statistical expertise on monitoring and advice on termination to a wide range of clinical investigators. Such a committee, perhaps under the sponsorship of a funding agency, would help to make the most efficient use of available research funds.

Vern T. Farewell is Professor and Richard J. Cook is Ph.D. student, Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.