18 November 1977, Volume 198, Number 4318 SCIENCE

Medical Research: Statistics and Ethics

Birnbaum Memorial Symposium

A one-day symposium entitled "Medical Research: Statistics and Ethics" was held on 27 May 1977 at the Memorial Sloan-Kettering Cancer Center in New York City, marking the opening of a new research facility, the Arnold and Marie Schwartz International Hall of Science for Cancer Research. The symposium was dedicated to the memory of Allan Birnbaum, an eminent statistician who until his death last year was associated with the establishment of a department of statistics at the medical center. The following articles are adapted from some of the lectures and discussions presented at the symposium.

Old Problems, New Challenges

Valerie Miké and Robert A. Good

The precept of Hippocrates (*1*) that "One must attend in medical practice not primarily to plausible theories, but to experience combined with reason" could have been written in our own day. Yet 2000 years were to pass before the scientific method became reality, and its fruitful application in medicine is an achievement of the 20th century.

The first large-scale randomized clinical trial was set up by the British Medical Research Council in 1946 to evaluate streptomycin in the treatment of pulmonary tuberculosis. This work was carried out under the leadership of Sir Austin Bradford Hill, who was largely responsible for introducing modern statistical concepts into medicine (*2*). An account of the historical development of therapeutic studies has been given by Bull (*3*). Today the design and analysis of clinical trials is an important area of statistical research.

The statistician newly engaged to exert his expertise in the field of medicine may have pondered the existing relation between experimental design, data analysis, statistical inference, and the scientific process. But usually he will not find himself adequately prepared for all the problems presented by his medical colleagues. He will encounter issues that bear little resemblance to what he has read in the pages of statistics journals or heard discussed by his academic associates.

He is to help design studies to determine the effects of different treatments on the course of human disease in the presence of other factors. The classical theory of statistical experimental design provides methods for the stimultaneous study of multiple factors and their interaction. Problems in agronomy had motivated Sir Ronald Fisher to develop such methods, and their impact was revolutionary. Advocating use of these techniques, Fisher wrote: "No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or, ideally, one question, at a time. [I am] convinced that this view is wholly mistaken. Nature . . . will best respond to a logical and carefully thought out questionnaire; indeed, if we ask her a single question, she will often refuse to answer until some other topic has been discussed" (*4*). When "levels" of factors may be controlled and replication is possible, statistical experimental design can be a powerful tool. The problem then is the extent of its applicability to the clinic.

The term "experiment" was itself introduced by Francis Bacon, who offered a scheme for generating observations and making inferences from them in a new system he called "experimental philosophy." Sir Peter Medawar, tracing the evolution of the concept of scientific method from Bacon to our own day, placed problems of design in science before those of validation. "It is a truism to say," he remarked, "that a 'good' experiment is precisely that which spares us the exertion of thinking: the better it is, the less we have to worry about its interpretation, about what it 'really' means" (*5*, p. 15). How near can we come to this laboratory ideal in the practice of medicine, where the terms of experiments perforce are set by nature and not by man? What does the statistician encounter in his daily work with medical investigators?

There is great need for investigating the natural history of disease, for analyzing the effects of past modalities of treatment and for conducting observational studies on essentially heterogeneous groups of patients. Much of this work is in the nature of exploratory analysis of very large data sets, a search for patterns providing clues to etiology or prognosis. We are here far removed from problems of mathematical modeling; we are studying complex phenomena whose mechanisms of action are not even vaguely understood. There is a tendency to place undue faith in the capabilities of the computer to provide the answers. We must bear in mind the subtleties of multiple testing, the fact that even random data will generally yield significant results. As Medawar has it: "We cannot browse over the field of nature like cows at pasture" (*5*, p. 29). And in this context the computer is super-cow. *P*-values are only flags here marking patterns of possible scientific significance, areas for further exploration on a selective basis. Exploratory data analysis is dependent on the insight of the subject-matter special-

Dr. Miké is head of the Biostatistics Laboratory of the Sloan-Kettering Institute for Cancer Research, New York 10021. Dr. Good is president and director of the Sloan-Kettering Institute for Cancer Research.

ist, in this case the clinical investigator, and his rapport with the statistician. He must convey to the statistician the essence of his intuitions.

What about prospective clinical trials? At our institution alone there are more than 200 currently active therapeutic trials, or "protocols." Even at a large hospital, the number of patients available for study in a relatively homogeneous group is generally small, while there may be several protocols competing for the same patient. Participating in the design of clinical studies, judgment must be exercised in finding a middle road between the impossible and the meaningless. At this stage, as well as at the time of analysis, we must emphasize the problem of the low power of our statistical tests, and the ambiguity of negative results in the presence of small sample sizes.

How do we handle the practice of repeatedly evaluating the progress of clinical trials, the problem again of multiplicity? The nature of the funding process imposes great pressure to publish. Requests for interim analysis of studies attain periodic peaks in our department on the days before abstract deadlines for the meetings of major medical societies. Such analysis is also required each time a progress report or renewal application is due. Everyone follows essentially the same procedure, using conventional significance levels for reporting results. Assessing the evidence presented at such meetings may be difficult if not impossible because of generally small sample sizes, and subtle to gross variations in protocols and types of patients studied. There is often heated debate centering on the sprinkling of statistically significant results. Studies are frequently terminated because a competing protocol yields good early returns. But in view of the enormous complexity of today's clinical medicine, and the sociology of our research structure, we wonder whether this situation can be changed.

Why is there such diversity of opinion on the pros and cons of randomized trials? Some believe that a trial is not ethical unless it is controlled, but at least one colleague feels that randomization in our setting is never ethical. There is no justification for adding to the anguish of the cancer patient by introducing the irrational element of "flipping a coin."

Medical ethics, the study of moral aspects of the physician-patient relationship, has fostered the development of codes of conduct for the medical profession. But interest in medical ethics is perhaps more widespread today than ever before. There is great public concern for human rights in the provision of medical care and the conduct of clinical research. New scientific, social, and legal developments need to be considered and reconciled. There are now institutes and scholarly journals devoted to the exploration of problems in "bioethics," and articles dealing with the subject also appear frequently in medical and other professional journals.

In addition to our own problems in clinical medicine, there is the spectrum of related issues currently in the public eye: the use of laetrile; the carcinogenicity of environmental agents; and the interrelationship of statistics, ethics, and public policy.

In a symposium entitled "Medical Research: Statistics and Ethics," conducted at the Memorial Sloan-Kettering Cancer Center, on 27 May 1977, several prominent members of the scientific community discussed some of these questions.

Tukey (*6*) gave an overview of statistical and ethical issues in the design and analysis of clinical trials, with special emphasis on problems of multiplicity. The subject was further developed by Mosteller (*7*), who used as starting point his analysis of published results of a series of randomized clinical trials dealing with innovations in surgery and anesthesia. Herbert (*8*) examined the ethical and legal aspects of human experimentation in general, including the complex issue of informed consent. Cornfield (*9*) addressed the problem of assessing the risk for humans of agents found to be carcinogenic in animals, touching on statistical, scientific, and public policy considerations. Cournand (*10*) discussed a proposed code of ethics for the scientist. Looking toward the future, he suggested the thesis that this code, complemented by an "ethic of development," could provide the blueprint for a worldwide scientific community as a source of humanizing influences to guide the development of man.

In a lecture at the symposium Dr. Archibald Cochrane, director of the Medical Research Council's Epidemiology Unit in South Wales, shared some of his personal recollections of how problems of medical ethics have been dealt with in Great Britain. A strong advocate of the use of controlled clinical trials in medical research, Cochrane is especially well known for his book *Effectiveness and Efficiency* (*11*), an appraisal of the British National Health Service.

In 1967, the Royal College of Physicians advised the setting up of Ethical Committees in all hospitals; evidence that they were badly needed was provided by the publication in the same year of *Human Guinea Pigs* by Pappworth (*12*). All research protocols must now be referred to these committees for approval, and the question naturally arises as to the extent of their effectiveness. Cochrane was asked to do a survey in 1969 in order to obtain some answers. The response rate to the survey was poor, so that his conclusions are of necessity based mostly on personal experience and private communication. An important salutary effect apparently has been the elimination, or marked reduction, of the types of research criticized by Pappworth, including undue experimentation involving prisoners and the mentally incompetent. There has also been a general improvement in the quality of research protocols. In Cochrane's opinion, however, the committees have been overprotective in preventing randomized controlled trials for the evaluation of established therapies. He also feels that the present procedure of complete secrecy is wrong and suggests that proposals judged unethical be published anonymously and thus be made available for general discussion. There is also the question of the attitudes of individual physicians. If the Ethical Committee has approved a protocol, should any physician still have the right not to participate, on the grounds that he deems the study unethical? This has been happening within the British system. Cochrane himself is firm in his own view that the randomized controlled trial is one of the most ethical forms of medical treatment.

What has the symposium accomplished? In the words of a colleague, the head of one of our clinical departments: "For those of us who stay largely occupied with the problems of clinical medicine, this was a timely reminder of the frail basis on which some of our decisions are made." Perhaps it will also have served as an invitation to continue the dialogue.

References

1. *Hippocrates*. W. H. S. Jones, Transl. (Heinemann, London, 1923), vol. 1, p. 313.
2. A. B. Hill, *Brit. Med. J.* **2**, 769 (1948); reprinted in *Statistical Methods in Clinical and Preventive Medicine*, (Oxford Univ. Press, New York, 1962).
3. J. P. Bull, *J. Chron. Dis.* **10**, 218 (1959).
4. R. A. Fisher, *J. Min. Agric. Gr. Br.* **33**, 503 (1926); reprinted in *Collected Papers of R. A. Fisher*, J. H. Bennett, Ed. (Univ. of Adelaide, Adelaide, 1971), vol. 2, p. 84.
5. P. B. Medawar, *Induction and Intuition in Scientific Thought* (American Philosophical Society, Philadelphia, 1969).
6. J. W. Tukey, *Science* **198**, 679 (1977).
7. J. P. Gilbert, B. McPeek, F. Mosteller, *ibid.*, p. 684.
8. V. Herbert, *ibid.*, p. 690.
9. J. Cornfield, *ibid.*, p. 693.
10. A. Cournand, *ibid.*, p. 699.
11. A. L. Cochrane, *Effectiveness and Efficiency* (Nuffield Provincial Hospitals Trust, London, 1972).
12. M. H. Pappworth, *Human Guinea Pigs* (Beacon Press, Boston, 1968).