STATISTICS IN CLINICAL RESEARCH

By D. D. Reid
London School of Hygiene and Tropical Medicine, London, England

All who have felt the intolerable itch to do research in any of the various fields of medicine must have met the feeling of frustration which so often accompanies one's first contact with the harsh realism of the statistical method. Nothing is more depressing than what has been called "the repellent symbolism" of the mathematical statistician. Yet, it is a comforting thought that no very sophisticated knowledge of the mathematical basis of statistical tests of significance is needed. What *is* essential for successful use of the statistical method in clinical research is a clear realization of the implied assumptions and the steps in the logical argument which are involved in the application of these mathematical procedures to clinical data. To make a parallel with therapeutic practice, none of us has ever felt inhibited in his use of a chemotherapeutic agent by an ignorance of the manner of its chemical synthesis. On the other hand, no one would dream of using a new drug whose dosage and possible toxic effects were unknown to him.

Statistical analysis can be a powerful weapon in the medical armory, but it must be wielded with insight and discrimination. The mathematical arguments involved in tests of statistical significance, for example, depend for their accuracy on the strict comparison of like with like as in treated and control groups of patients. They will do nothing to eradicate any basic faults in the collection of the original data. Statistical analysis, therefore, should *not* be used rather as an afterthought at the "post-mortem" or autopsy of an experiment. Statistical reasoning is needed as soon as that experiment is conceived in the mind of the research worker and throughout its conduct. As Fisher remarked: "The statistician must be treated less like a conjurer whose business it is to exceed expectation, than as a chemist who undertakes to assay how much of value the material submitted to him contains." Indeed, if you feel the need for technical assistance, you should consider your statistical colleague rather as an architect, to be consulted *before* the work is started, so that, by taking thought together, both experimenter and statistician can insure that the material will be collected in such a way as to give the maximum amount of accurate information. Only thus can the clinical research worker escape the vitriol of our comment. Only thus can we medical statisticians escape the ignominious label of armchair critic.

The descriptions in previous papers of the various experimental designs which are applicable in such fields of biology as agriculture are fascinating in their subtle ingenuity. I think most medical statisticians have, at one time or another in their career, been mesmerized by the possibilities opened up by their use in clinical trials. Unfortunately, it has been our experience that, for a variety of reasons, ethical as well as practical, the precise manipulation of treatment required by the more complex experimental designs

has not been possible. Nevertheless, we believe that much can be achieved by the use in clinical trials of simple robust methods which observe the basic principles outlined by Dr. Mainland. I propose, therefore, in the remaining part of this paper, to outline some of the statistical planning aspects of the clinical trials of streptomycin in pulmonary tuberculosis recently conducted in Britain[1]. The report of that trial expresses very effectively the trend of our thinking on this subject in Britain at the moment.

As previous experience with the use of gold therapy in pulmonary tuberculosis had shown, the vagaries of the natural course of that disease made uncontrolled trials quite unreliable. On the other hand, early optimistic reports from the United States about the value of streptomycin had created considerable pressure for its immediate adoption in the routine treatment of tuberculosis. In the meningeal form of the disease, when the case fatality rate is practically 100 per cent, controls were, of course, quite unnecessary. Whatever ill effects the drug might have, it could not increase the patient's chance of death beyond that 100 per cent. No ethical or moral problems were involved in running uncontrolled trials in tuberculous meningitis.

In pulmonary tuberculosis, on the other hand, ethical questions arise. The denial to a seriously ill patient of a drug which appeared to be of great value in this disease was involved in his inclusion in the untreated or control group of patients. The possibility of such an added risk to that individual was a consideration which hampered the initiation of controlled trials in America, and, until 1946 at least, no such trials had been reported in America.

In Britain, however, our temporary financial embarrassment limited the supply of streptomycin, and a heavy expenditure of hard currency on streptomycin for pulmonary tuberculosis could be justified only after an adequate trial of its capabilities. Our genteel poverty thus paid a scientific dividend by quieting any doubts we might have about the ethics of controlled trials of streptomycin in pulmonary tuberculosis. In short, where only a few could be treated, many would have to remain untreated in the normal course of events. It was decided, therefore, to make the best of this situation by running a well-controlled trial of streptomycin in pulmonary tuberculosis.

As a *first* step—and I would stress the word, first—a co-operative working group of clinicians, pathologists, radiologists, *and* a statistician, Professor Bradford Hill, was formed to make the experimental plan, and a co-ordinating and supervising research worker was appointed to insure its competent execution.

For several reasons, this group decided to restrict the trial to pulmonary tuberculosis of a clearly defined type. In fact, the trials included only patients aged between 15 and 30 suffering from progressive bilateral pulmonary tuberculosis of presumably recent origin, bacteriologically proved and unsuitable for collapse therapy. Such a rigid definition was useful in that it insured the homogeneity of the experimental group. On the other hand, the acceptance of such restrictions in the sample of cases of pulmonary tuberculosis meant the acceptance of a similar restriction in the generality of the results achieved. That is to say, the experience noted in the trial could only be applied generally to cases of the type included in the trial.

This restriction was acceptable for two reasons: the homogeneity of the sample minimized the number of sources of variation between individuals which might have obscured the main issue; and, secondly, the fact that such cases would normally have been treated by rest in bed alone simplified the ethical problem of control.

A notable feature of this trial was the frank realization by all concerned of the fallibility of human judgment in general and of clinical and radiological judgment in particular. At all stages of the trial, then, precise criteria of diagnosis, progress, and cure were laid down, and all judgments on X-ray findings were made by two or more observers, independently of each other and unbiased by any knowledge of the nature of the treatment given to the patient whose physical status was being assessed.

This principle of the elimination of personal bias is fundamental in all experiment, but it is of particular importance in clinical research. Thus, in the selection of patients for inclusion in either treated or control groups, the final decision was made purely on a chance basis. Once a panel of clinicians had decided on the suitability of a case for inclusion in the trials, the mode of treatment was decided quite independently by reference to a series of cards enclosed in sealed envelopes bearing the name of the hospital to which the patient was to be sent and the order of his arrival there. On the card had been inscribed a letter S or C, which denoted the group to which the patient should be allocated. These letters had been previously decided by Professor Hill, in an absolutely random manner, by reference to a list of what are called random sampling numbers. The clinicians on the panel were thus absolved from any responsibility in the allocation of patients, and the possible effects of personal bias in selecting cases for treatment were rigorously excluded. It is always well, even with such a procedure, to check, by comparing their relevant characteristics such as age and condition on admission, that the two groups thus arbitrarily selected are in fact alike.

As the treatment of these patients went on, the results of objective tests of temperature ranges and sedimentation rates were recorded in a standardized manner, and routine X-ray checks were made by independent observers, unbiased by any knowledge of the patients' identity or mode of treatment. In the final assessment again, there recurs this insistent theme of the elimination of personal bias, which is essential for success in clinical research.

The final statistical analysis by the application of the χ^2 and t tests (which conclusively demonstrated the beneficial effects of streptomycin) could thus be confidently made, since we were reasonably sure that the rain of chance events had fallen equally upon the just and the unjust. The differences observed, *e.g.*, in case-fatality rates, between treated and control groups may be due to chance, and it is the function of the technical test of significance to test just that hypothesis. Indeed, it is as well to reverse the normal processes of Anglo-American justice. Chance is always considered to be guilty or responsible for the differences until its innocence has been proved by the results of technical tests of significance. Then, and in general only then, can alternative explanations for these differences be considered.

Last among these alternatives, you should consider the possibility that your own therapeutic brain-child was really producing a beneficial effect.

The discussion in the streptomycin report is an excellent example of the soul-searching which is so essential a part of the interpretation of statistical analysis. One must always be on the look-out for a confusion between the demonstration of a mathematical relationship and the proof of cause and effect. The association between differences in treatment and differences in outcome does not necessarily mean that the treatment has improved prognosis. The improvement may well result from some unforeseen and uncontrolled source of variation which had a selective action on one of the groups. In the present instance, for example, the onset of spontaneous pneumothorax (which occurred slightly more frequently in the streptomycin-treated group) might well have obscured the final result. One must not, therefore, allow one's regard for the statistical coefficient to outweigh one's knowledge of the subject matter. If the result of your statistical analysis does not agree with logical expectation, then the least you should do is check your arithmetic. On the other hand, it would be wrong to confuse common sense with personal preconceptions and prejudices. If the experimental plan has been soundly conceived and executed and the results clash with your notions about the eternal verities of clinical medicine, then, like Cromwell, "I beseech you, gentlemen, consider it possible that you may be mistaken." Statistical methods may be no substitute for common sense, but they are often a powerful aid to it.

In the final assessment of the results of clinical trials, I can do no better than repeat one version of the Queen's advice to Alice:

> "Don't go on like that!" cried the Queen,
> "Consider what a great girl you are.
> Consider what a long way you've come.
> Consider what o'clock it is.
> Consider everything."

For to "consider everything" is the keynote of success in the interpretation of the results of clinical experimentation.

Reference

1. Medical Research Council. 1948. Streptomycin treatment of pulmonary tuberculosis. Brit. Med. J. **ii**: 769.