THE STATISTICIAN AND MEDICAL RESEARCH

BY

MAJOR GREENWOOD, D.Sc., F.R.C.P., F.R.S.

Professor Emeritus of Epidemiology and Vital Statistics in the University of London

Statistics is—or are—an old science (or art), but Francis Galton, Karl Pearson, and W. F. R. Weldon discovered a new application of statistical method at the end of the nineteenth century, and in 1902 the biometricians, as they were called, established a journal of their own, *Biometrika*, now in its 35th volume. The first volume of *Biometrika* contains a paper of medical interest—on the influence of previous vaccination in cases of smallpox—and others followed. For some years, however, the influence of the new race of statisticians on the logical procedures of either pathologists or clinicians was small and largely confined to destructive criticism. Biometrician A.B.—the actual letters were usually K.P.—would fall upon a paper by Dr. C.D. and show that Dr. C.D.'s conclusion that treatment X gave better results than treatment Y did not follow from his data, that there was no significant (then and now a terribly overworked word) difference between the means of his two series. Even when, as sometimes happened, the difference was significant, A.B. would argue that other differences between the two series than the method of treatment might account for the result.

At first both laboratory workers and clinicians simply ignored the biometricians or called them mere mathematicians, often thought by clinicians and some biologists, like Almroth Wright, to be an inferior race of mankind. As J. J. Thomson once said, if ignorance of mathematics is a virtue it is surely the easiest of virtues. However, the tide turned; in 1910 the Lister Institute appointed a medical statistician and the Medical Research Committee (now Council) appointed another in 1914. Many statisticians now well known would acknowledge how much they owe to the friendship and encouragement of Sir Charles Martin or Sir Walter Fletcher which brought them through times of depression and frustration.

The widespread activities of the Medical Research Council and of an associated research committee, now the Industrial Health Research Board, led to increasing utilization of statistical methods, even in clinical research. In 1921 a statistical research committee was formed to take part in the planning and control of industrial-medical inquiries. Four years later this committee's terms of reference were extended and it became the statistical committee of the Medical Research Council itself. In the succeeding 14 years this committee had a large part in bringing statistical methods to bear upon laboratory researches as well as upon those field or factory inquiries which admittedly involved statistical analysis.

Contemporaneously, new statistical methods were used in non-medical laboratory and field biology. Progress in this way in agricultural research was particularly great, because one of the most gifted of contemporary statisticians, Professor R. A. Fisher, was for several years head of the statistical department of the Rothamsted Experimental Station. His epoch-making additions to statistical technique, applicable far beyond the bounds of soil research, were naturally first used in agricultural trials.

Now the biometricians have triumphed; indeed, some people may be reminded of Macaulay's rhetoric, that the whigs endured everything that O'Connell should not "be less than a British subject. We never will suffer him to be more."

Demand for Statisticians

One may say, broadly, that the demand for statisticians has increased for two reasons. The first is that statistical methods have become far more exact and searching—i.e., capable of reaching results in fields which even so recently as 50 years ago seemed closed to them. The next is that experiments are expensive in money and time; we want results as cheaply as possible and as quickly as possible. It is therefore important to squeeze the last drop of information out of an experiment—i.e., to plan it properly and to analyse the data in the most efficient way. Suppose the experiment is a clinical one, that we desire to test the values of two treatments—A and B. I think, before Professor Fisher taught us 'better, that we should have looked at it in the following way. If we had available, say, 120 patients we should give A to 40, B to 40, and leave another 40 untreated.

Our 120 patients are not test-tubes, they differ in many ways; some of these differences may be relevant to the success of treatment, hence one must have a technique for the assignment of patients to each of the three groups which does not weight one group unfairly with persons likely to respond more (or less) favourably than others to treatment. But a statistician would not now recommend three groups of 40, unless remedies A and B were known to be incompatible. If, for example, the treatments were two *different* techniques for removing a new growth, clearly to give a group of patients both treatments is nonsensical. But often A and B can be combined; if therefore we had four sets of three: (1) AB, (2) A, (3) B, (4) neither, and assigned 30 patients to each, then not only should we be able to test whether the combination of A with B were better than either, which the "pure" experiment did not permit, but we have lost nothing in stringency of comparison of single treatments, for we can not only compare A against the "pure" controls but AB against B, if we wish to assess the value of A alone. This is a simple example of the economy of factorial design, and illustrates the fallacy of isolating factors, a doctrine which, as Professor R. A. Fisher remarked, "seems to be more nearly related to expositions of elementary physical theory than to laboratory practice in any branch of research."

The New Method

An example of the power of statistical methods in medical research is the assaying of drugs by biological tests. Up to comparatively recent times laboratory workers were apt to look on mice or guinea-pigs just as test-tubes of different sizes and ages. I think the first worker in England to allow statistically for the fact that, just as mice individually vary in weight, appetite, and temper, they vary in response to doses of toxins or drugs, was J. W. Trevan, who published an important paper in 1927. In 1933 J. H. Gaddum's Report to the Medical Research Council extended the new method, and between 1937 and 1943 J. O. Irwin, of the Medical Research Council's statistical staff, published several valuable papers. In America, from 1934 onwards, C. I. Bliss and his colleagues did much research into the theory and practice of the method.

D. J. Finney's recent book (*Probit Analysis*, Cambridge University Press, 1947) gives a clear account of this new and valuable technique. The fundamental idea is simple enough. We are interested in a drug which can be tested only by administering doses of it to samples of mice and noting how many die at each dosage. Common sense dictates the drawing of a graph of the percentages which die at each dosage plotted as ordinates against the dose, or some function of the dose, as abscissa. But we must take account of the variability of response and remember

that even if we use equal numbers of mice at each dosage the points on the graph are not equally well determined, because, for statistical reasons, the precision of the ordinate for 50% fatalities will be greater than for 10% or 90%. Trevan noted that taking for abscissae the logarithms of the doses gave better results than using the actual doses, and a further transformation reduced the problem to the fitting of a straight line; now, very efficient routine methods of assay are available. But to reach this has involved a good deal of mathematical-statistical analysis.

Lastly, statistical methods play an important part in applied psychology. The pioneer was Charles Spearman, who concluded that scores in intellectual tests—for instance, examination marks in French or chemistry—could be expressed in terms of two “factors”—a general factor measuring something common to all cognitive activities and a special factor peculiar to each activity. Later research has shown the problem to be more complex than it first seemed to be, but psychological testing has grown and extended from the cognitive to the conative aspects of the mind and therefore interests psychiatrists and medical psychologists as well as schoolmasters and Civil Service examiners. Sir Cyril Burt’s treatise, *The Factors of the Mind* (University of London Press, Ltd., London, 1940), gives a critical review of what is now called “factor-analysis.” Here again the construction of an adequate arithmetical routine involved a good deal of difficult mathematical analysis.

I have said nothing of the importance of the simple statistical methods, already understood well enough 50 years ago, in clinical medicine. Professor J. A. Ryle in his recent book (see *Changing Disciplines*, particularly p. 104 *et seq.*, Oxford Medical Publications, 1948) has spoken of this with a clinical authority I lack.

B.M.A. Committee’s Report

Perhaps the reader is satisfied of the importance of statistical methods in all medical research, and he is sure to know that in the present undergraduate medical curriculum statistics, like liturgy, civil engineering, or the use of the globes, has no place at all. Even from the professional point of view this is a pity; my quotation from Macaulay’s speech was not made at random. The medically trained person, whether laboratory worker or clinician, can no longer afford to despise the statistician, who for many years to come will *not* be medically qualified. He must accept him as a colleague, and unless the “doctor” does have some knowledge of the elements of statistical method the colleague will become, in the eyes of the world—which is becoming statistical-minded—not the colleague but the leader. Yet, down to the publication of the B.M.A. Curriculum Committee’s report on *The Training of a Doctor* (Butterworth’s Medical Publications, 1948), no authoritative body has faced the position.

This report differs from those published by the Stationery Office during the war on educational problems, medical or general, in two respects; it is written in the English language, not in jargon, and can be read with pleasure because it develops a plan logically and artistically. The committee thinks that statistical method should have a place in the first year of medical study because statistical ideas have their part in shaping current views of the fundamental sciences—namely, physics, biology, and chemistry—it recommends a course of 10 lectures, with practical work, and sets out a very reasonable syllabus. The whole of the committee’s remarks (paras. 86–91 on pp. 28–9) seem to me excellent. If I criticize the proposal it is not from lack of sympathy with the enlightened spirit which inspired it. A

class of first-year medical students will consist of young men and women who for the most part have learned *and remembered* little more mathematics than what is covered by the ordinary school-leaving certificate; indeed, if some medical faculty boards have their way, and medical students are not required to take mathematics even up to this standard, the average mathematical level of the class may be even lower. It is quite true that the principles of statistical method can be taught, and efficiently taught, without any more technical knowledge of algebra than many of the class will have or could quickly acquire. But that requires more time and a far more efficient and experienced teacher than the instruction of pupils with a serious mathematical background.

I do not think formal lectures should bulk in the course of instruction. No doubt an introductory lecture might rouse the enthusiasm of the class, be good propaganda, if the teacher had the gift of persuasiveness; it might, however, be merely dull. I should say that the best way of teaching is informal; one should try to combine a demonstration with a practical class. The teacher might, for example, give a short talk, say 10 minutes, on the advantages and disadvantages of common averages, the mean, the median, and the mode, and how these are computed. The class should then follow the heuristic method of Mr. Squeers and at once work out examples for themselves. As every experienced speaker or teacher knows, “impromptu” speaking and “informal” teaching require much more preparation than a set speech or formal lecture. In our particular case, worked examples and instructions to students can be mimeographed and distributed, but the difficulties many intelligent people have with “sums” are infinite, and the teacher must have had a good deal of experience to be able to remove or alleviate the student’s perplexities. I am a pure Cockney, and my experience as student and teacher is wholly confined to London. In London there are not now enough teachers of the necessary powers to go round the medical schools. This might not be an unmixed evil. Although a statistical laboratory is not so expensive to equip as a physiological laboratory, aids to computation, such as calculating machines, cost a good deal of money. To confine practical work in statistics to what can be done by brute force and a table of logarithms would be as wise (or as foolish) as to restrict experimental physiology to what could be done on oneself—or a fellow student—with one’s wits and ten fingers. Two or three equipped departments might receive all the medical students in London, in a regular rotation so that the classes were small:

But unless the committee can persuade members of the faculty board and of the boards of studies within the faculty of medicine that statistics is not an elegant extra but a fundamental discipline, not indeed demanding as much time as chemistry, physics, and biology, but still fundamental, its plan will not succeed. Nobody dislikes examinations more than I do, but unless “human nature” has changed a great deal in the last 50 years—perhaps it has—the average student will not take much trouble over subjects unremunerative in marks. I hope the committee will succeed; all its members are clinicians and, for sound psychological reasons, their opinions are weighty. I think I should find it as difficult to persuade the board of the faculty of medicine to take the committee’s plan seriously as a cobbler would find it to persuade me to double my expenditure on boots.

During 1947, 60% of fatal accidents and 38% of all accidents on building work reported to the Factory Inspectorate were on jobs not governed by the existing regulations for the safety and welfare of the workers. The Minister of Labour has therefore made new regulations under the Factories Act, 1937, to cover such cases.