PRINCIPLES OF MEDICAL STATISTICS

I.—THE AIM OF THE STATISTICAL METHOD

"Is the application of the numerical method to the subject matter of medicine a trivial and time-wasting ingenuity as some hold, or is it an important stage in the development of our art, as others proclaim?"

WHATEVER may be our reactions to that question, propounded by Prof. Major Greenwood in these columns[1] fifteen years ago, it must be admitted that in the years intervening between 1921 and to-day there has been a substantial increase in the number of papers contributed to medical journals of which the essence is largely statistical. Not only does there appear to be an enhanced knowledge of and respect for the national registers of life and death which the Registrar-Generals of the United Kingdom annually publish and analyse, but there is an increasing number of workers who endeavour to apply numerical methods of analysis to their records obtained in clinical medicine. The great majority of such workers, however, have had, naturally enough, little or no training in statistical method and many of them would find the more mathematical methods of the professional statistician, as has been said, "obscure and even repellent." Often enough, indeed, the argument is put forward that the use of such mathematical methods is quite unjustifiable, that the accuracy of the original material is not sufficient to bear the weight of the treatment meted out to it. This assertion is not strictly logical. If a collection of figures is worth a statistical analysis at all, it is, obviously, worth the best form of statistical analysis—i.e., the form which allows the maximum amount of information to be derived from the data.

Whether mathematical statistical methods *are* the best form in particular cases or may be regarded as an unnecessary elaboration must turn rather upon this question: Can we in any of the problems of medical statistics reach satisfactory results by means of relatively simple numerical methods only? Or upon this question: Can we satisfactorily test hypotheses and draw deductions from data that have been analysed by means of such simple methods? Personally I believe the answer to these questions is an unqualified yes, that many of the figures included in medical papers can by relatively simple statistical methods be made to yield information of value, that where the yield is less than that which might be obtained by more erudite methods which are not at the worker's command the best should not be made the enemy of the good, and that even the simplest statistical analysis carried out logically and carefully is an aid to clear thinking with regard to the meaning and limitations of the original records. If these conclusions are accepted, the question immediately at issue becomes this: Are simple methods of the interpretation of figures only a synonym for common sense or do they involve an art or knowledge which can be imparted? Familiarity with medical statistics leads inevitably to the conclusion that common sense is *not* enough. Mistakes which when pointed out look extremely foolish are quite frequently made by intelligent persons, and the same mistakes, or types of mistakes, crop up again and again. There is often lacking what has been called a "statistical tact, which is rather more than simple good sense." That tact the majority of persons must acquire (with a minority it is undoubtedly innate) by a study of the basic principles of statistical method.

It is my object in these articles to discuss these basic principles in an elementary way and to point out by representative examples taken from medical literature how these principles are frequently forgotten or ignored. I very much fear that the discussion will often appear too simple and that some of the mistakes to which space is given will be thought too futile to need attention. I am only persuaded that such is not the case by the recurrence of these mistakes and the neglect of these elementary principles, a feature with which every professional statistician is familiar in the papers submitted to him by their authors for "counsel's opinion."

Definition of Statistics

Whereas the laboratory worker can frequently exclude variables in which he is not interested and confine his attention to one or more controlled factors at a time, the clinician and social worker have to use records which they know may be influenced by factors which they cannot control but have essentially to be taken into account. The essence of the statistical method lies in the elucidation of the effects of these multiple causes. By statistics, therefore, we mean "quantitative data affected to a marked extent by a multiplicity of causes," and by statistical method "methods specially adapted to the elucidation of quantitative data affected by a multiplicity of causes" (G. U. Yule: An Introduction to the Theory of Statistics, 1927). For example, suppose we have a number of children all of whom have been in contact with measles and to a proportion of them is given an injection of convalescent serum. We wish to know whether the treatment prevents the development of a clinical attack. It is possible that the risk of developing an attack is influenced by age, by sex, by social class and all that that denotes, by duration and intimacy of contact, by general state of health. A statistical analysis necessitates attention to *all* these possible influences. We must endeavour to equalise the groups we compare in every possibly influential respect except in the one factor at issue—namely, serum treatment. If we have been unable to equalise the groups ab initio we must equalise them to the utmost extent by the mode of analysis. As far as possible it is clear, however, that we should endeavour to eliminate, or allow for, these extraneous or disturbing causes when the experiment is planned; in a carefully planned experiment we may determine not only whether serum is of value but whether it is more efficacious at one age than another, &c. It is a serious mistake to rely upon the statistical method to eliminate disturbing factors at the completion of the work. *No* statistical method can compensate for a badly planned experiment.

Planning and Interpretation of Experiments

It follows that the statistician may be able to advise upon the statistical lines an experiment such as that referred to above should follow. Elaborate experiments can be planned in which quite a number of factors can be taken into account statistically at the same time (R. A. Fisher: The Design of Experiments, 1935). It is not possible to discuss such methods here and attention is confined to the type of simple experimental lay out with which medical workers are familiar. Limitation of the discussion to that type must not be taken to mean

[1] Medical Statistics (1921) *Lancet*, **1**, 985.

that it is the best form of experiment in a particular case.

The essence of the problem in a simple experiment is, as emphasised above, to ensure beforehand that, as far as is possible, the control and treated groups are the same in all *relevant* respects. The word relevant needs emphasis for two reasons. First, it is obvious that no statistician can be aware of all the factors that are, or may be, relevant in particular medical problems. From general experience he may be able to suggest certain broad disturbing causes which should be considered in planning the experiment (such as age and sex in the example above) but with factors which are narrowly specific to a particular problem he cannot be expected to be familiar. The onus of knowing what is likely to be relevant in a specific problem must rest upon the experimenter who is, presumably, familiar with that narrow field. Thus, when the statistician's help is required it is his task to suggest means of allowing for the disturbing causes, either in planning the experiment or in analysing the results, and not, as a rule, to determine what *are* the relevant disturbing causes.

The second point that must be observed as regards the equality of groups in all relevant respects is the caution that must attend the interpretation of statistical results. If we find that Group A differs from Group B in some characteristic, say, its mortality-rate, can we be certain that that difference is due to the fact that Group A was inoculated (for example) and Group B was uninoculated? Are we certain that Group A does not differ from Group B in some other character relevant to the issues as well as in the presence or absence of inoculation? For instance, in a particular case, inoculated persons might, on the average, belong to a higher social class than the uninoculated and therefore live in surroundings in which the risk of infection was less. We can never be *certain* that we have not overlooked some relevant factor or that some factor is not present which could not be foreseen or identified. It is because he knows a complex chain of causation is so often involved that the statistician is, as it appears to many persons, an unduly cautious and sceptical individual.

The reason why in experiments in the treatment of disease the allocation of alternate cases to the treated and untreated groups is often satisfactory, is because no conscious or unconscious bias can enter in, as it may in any selection of cases, and because *in the long run* we can fairly rely upon this random allotment of the patients to equalise in the two groups the distribution of other characteristics that may be important. Between the individuals within each group there will often be wide differences in characteristics, for instance in body-weight and state of health, but with *large* numbers we can be reasonably sure that the numbers of each type will be equally, or nearly equally, represented in both groups. If it be known that certain characteristics will have an influence upon the results of treatment and on account of relatively small numbers the distribution of these characteristics may not be equalised in the final groups, it is advisable to extend this method of allocation. For instance, alternate persons will not be treated but a division will be made by sex, so that the first male is treated and the second male untreated, the first female is treated and the second female untreated. Similarly age may be equalised by treating alternate males and alternate females at each age, or in each broad age-group if individuals whose ages are within a few years of one another may in the particular case be regarded as equivalent.

AN EXAMPLE

Even if in planning the experiment factors such as age and sex have thus been equalised in the two groups, the analysis of the results should never be directed only to the effects of treatment upon the totals. It is possible that treatment may be more effective at one age than at another and this difference will be lost sight of if the groups as a whole are alone compared. As far as the number of observations allow, comparison should be made between sub-groups differentiated by any factors that may be of importance. As an example of this experimental method and analysis may be cited the report of the Therapeutic Trials Committee of the Medical Research Council on the serum treatment of lobar pneumonia (*Brit. med. J.* 1934, **1**, 241). In each of three centres, London, Edinburgh, and Aberdeen, it was arranged that alternate cases taken in order of their admission to hospital should be treated with serum, both treated and untreated being so far as possible in the same wards and under the care of the same physicians. "It was thought better not to attempt a deliberate sorting of cases in respect of mildness or severity, but to trust that the distortion of chance scatter would become almost negligible in a fairly large number of cases." Age, it will be observed, was not equalised in the allotment of cases, and as fatality from lobar pneumonia certainly varies with age the analysis must first show whether the serum and control groups are equivalent in that respect. The random allocation of patients has, in fact, not quite equalised the groups in that factor. Taking the figures for London and Aberdeen (in Edinburgh the method was not fully carried out) and combining Types I and II, the numbers in two broad age-groups were:—

Age.	Number of patients.		Percentage in each age-group.	
	Controls.	Serum treated.	Controls.	Serum treated.
20–40 ..	104	123	65	75
40–60 ..	55	40	35	25
Total ..	159	163	100	100

Roughly 35 per cent. of the control cases were of ages 40–60 and only 25 per cent. of the serum-treated. It follows that the comparison of results must essentially be made within, at least, these two age-groups. If a lower fatality-rate is found in the serum-treated when the total groups are compared, this result may merely be a reflection of the fact that 10 per cent. less of these patients belonged to the ages at which fatality is normally higher and, correspondingly, 10 per cent. more belonged to the ages at which fatality is normally lower. In other words, as the method of allotment has not equalised the groups in an important respect with regard to fatality, the method of analysis must do so.

It is impossible to ignore the fact that in the random allocation of patients to the treated and untreated categories a difficult moral issue is often raised. The treatment is usually based on a priori evidence which suggests that it should have some curative effect. Can it, then, be justifiably withheld from any patient? And if it is withheld how extensive a trial is justifiable? There are no easy answers to these

questions. There can be no doubt that any new therapeutic measure *should* be given a period of trial before coming into general use. If the results of such a measure are dramatic in the light of all past experience, then clearly the trial period will not need to be prolonged. But one must be careful of the interpretation of "dramatic"; it must imply something strikingly different from all previous experience, not merely a run of a few successes where some failures might well have been expected. Such dramatic events are unfortunately the exception. More usually the effect of the new measure will be relatively small, though none the less important, and therefore its presence (or absence) be difficult to detect in small-scale tests. Such tests may well give contradictory and confusing answers. They may enable a useless measure to hold for years a position it does not merit (sometimes, perhaps, to the exclusion of more worthy measures); they may even prevent a really valuable form of treatment obtaining the general recognition it deserves. A well-planned and extensive trial (such as has been carried out with measles serum) would obviate such undesirable results, and thereby possibly save more suffering in the long run than is incurred in the trial itself, and more speedily than the indeterminate results given by inadequate tests would allow.

Summary

The statistical method is required in the interpretation of figures which are at the mercy of numerous influences, and its object is to determine whether individual influences can be isolated and their effects measured. The essence of the method lies in the determination that we are really comparing like with like, and that we have not overlooked a relevant factor which is present in Group A and absent from Group B. In experiments involving the treatment of a number of patients who are to be compared with controls not given the specific treatment, any deliberate choice of individuals to be treated may lead, unconsciously, to the treated group differing from the untreated group in some characteristic which, known or unknown, has an influence upon the results. If the series of cases is large, a random allocation of individuals—e.g., by cases being alternately placed in the treated and untreated groups—may reasonably be relied upon to equalise the two groups in all characteristics except the one under examination. Such a method of allocation does not obviate the necessity of testing in the statistical analysis of the results whether in fact the groups are equal in all characteristics which are believed to be relevant.

A. BRADFORD HILL.

SPECIAL ARTICLES

INCIDENCE OF ANÆMIA IN PREGNANCY

INFLUENCE OF SOCIAL CIRCUMSTANCES AND OTHER FACTORS

BY W. J. S. REID, B.Sc., M.D. Aberd., M.R.C.P. Lond.
HON. PHYSICIAN TO ANCOATS HOSPITAL, MANCHESTER; AND

JEAN M. MACKINTOSH, M.D. Glasg., D.P.H.
ASSISTANT MEDICAL OFFICER FOR MATERNITY AND CHILD WELFARE, COUNTY BOROUGH OF STOCKPORT

A SERIES of 1108 pregnant women was examined. There was no selection of cases either as to social circumstances or as to parity, the patients being those who attended the Stockport corporation antenatal clinics. The Haldane hæmoglobinometer was used, and the hæmoglobin estimations were all made by one of us (J. M. M.). The examinations extended over about two years, and were more or less evenly distributed throughout the seasons. About half the women were examined in the forenoon and the rest in the afternoon. Some 72 per cent. of the examinations were made between the twenty-sixth and the thirty-second week of pregnancy, and only 7 per cent. after the thirty-sixth week. The age and parity distribution of the women is shown in Table I.

Incidence of Hyperchromic Anæmia.—In all cases showing a reading of 70 per cent. of hæmoglobin and under, blood films were prepared and examined by one of us (W. J. S. R.). No case of pernicious anæmia was found. It may be recalled that Evans, who examined 4083 patients, "found no case of severe anæmia."

Incidence of Hypochromic Anæmia.—Amongst our cases we found that 45·9 per cent. gave a hæmoglobin reading of 86 per cent. and over, 43·9 per cent. showed a reading between 70 and 84 per cent., and 10·2 per cent. gave a reading below 70 per cent. Davidson and his colleagues in Aberdeen (1935) in an examination of 819 pregnant women, found 17·5 per cent. with a hæmoglobin reading below 70 per cent. Helen Mackay (1935) in London examined 109 pregnant women and found 4·6 per cent. with a hæmoglobin reading below 70. Boycott (1936), also working in London, in a series of 222 cases found 11 per cent. of his cases gave a reading below 70. All these workers used the Haldane method. American investigators, Moore (1929), Lyon (1929), Galloway (1929), Adair, Bland, Goldstein, First (1930), with various methods of estimation, examined groups of cases ranging from 100 to 1176 in number, and found a much greater incidence of anæmia in pregnancy. They give figures ranging from 19 to 81 per cent. of patients with less than 70 per cent. hæmoglobin.

TABLE I

Age-group.	Parity. 1	2	3	4	5	6	7	8	9	10	11	12	13	Total
20 and under.	75	14	2	—	—	—	—	—	—	—	—	—	—	91
21–25	205	92	27	12	2	1	—	—	—	—	—	—	—	339
26–30	117	99	79	25	18	11	—	—	—	—	—	—	—	349
31–35	31	40	29	36	28	10	12	7	4	1	1	—	—	199
36–40	8	19	6	14	11	11	8	3	6	3	3	4	—	96
41 and over.	4	1	5	2	6	4	3	2	1	3	2	—	1	34
Total	440	265	148	89	65	37	23	12	11	7	6	4	1	1108

Effect of Social Circumstances.—In order to determine whether economic circumstances had any effect, we divided our patients into two groups, according to the average income per head of the family per week after deduction of rent. Group I. comprised those women whose family income amounted to an average of 12*s.* per head per week or less; Group II. those with an average over 12*s.* per head. It was possible to assess the income in each case with a considerable degree of accuracy from office