A COMMENT ON SOME ASPECTS OF MEDICAL STATISTICS*

By Tye Cho-Yook, B.A.

(Department of Social Medicine & Public Health, University of Malaya in Singapore)

The growth of Statistics is attributable to the increasing recognition of the need to study the behaviour patterns and characteristics of groups of persons collectively as opposed to individual persons. Clinical medical practice is essentially the treatment of each individual as a case peculiar to itself. This is tacit recognition of the fact that there is variation between individuals, not only in their presenting characteristics or symptoms, but also in their response to treatment. In short, as the saying goes, no two persons are alike. For a long time people used this as an argument against the usefulness of studies made on groups of subjects collectively, and some even went so far as to say that statistics had no place in clinical medicine. However, statistics thrive on variations. As a matter of fact, it is the very existence of variations that makes the statistical approach so necessary for the study of human problems. For if there were no variation—if everyone were alike in every way concerned—there would indeed be no place for statistics. By studying one individual we would have studied the lot and there would be no need whatsoever to study groups of persons collectively.

There is a recognizable and consistent pattern in the nature of variations of all types, and these variations exist only because of the countless different ways in which a multiplicity of factors can affect each individual. But a community of individuals collectively would continue to manifest the same consistent pattern so long as the relevant factors continued to affect it to the same degree. To give an example: The risk of dying of an individual depends on the presence or absence of disease, the nature and duration of disease and whether he seeks treatment when he is sick; whether he takes care when he crosses a busy street; whether he is partial to driving a sports car recklessly at 100 miles per hour; whether he overeats or undereats and what he eats or drinks; whether he was vaccinated; whether he works in a mine underground or sits at a desk in an air-conditioned office; whether he is young or old, male or female, married or unmarried, whether anyone is thinking of assassinating him, and a whole host of other known and unknown, measurable and unmeasurable factors. We can appreciate the multiplicity of factors concerned, medical, sociological, psychological, genetic, environmental, etc. Yet the proportion of people dying (the death rate) in the community as a whole, will remain very much the same each year so long as the sum total of influence of these factors remains unchanged in that community. In other words, we can predict the approximate number of people who will die, but we can't predict their identities. Statistical procedures, however, enable us to classify the population concerned into various relevant categories, and to analyse the differential risks of dying for each sub-group in relation to the differential factors affecting the individuals in it, thereby evaluating quantitatively and objectively the role of each of the factors in turn.

We can postulate any number of other examples such as the risk of getting a particular disease or the prevalence of dental caries or the chance of recovery following administration of a particular treatment to patients suffering from a particular disease. The same line of argument is applicable. Whatever the condition we may be examining in our respective fields of interest, the condition is always the result of interaction between a variety of factors.

In both preventive and curative medicine, we are concerned with the need to discover and identify as many as we can of the more important factors which determine the health of individuals, and to modify, remove, replace, or control them in such a way as to restore the individual to and maintain him in good health. This is the essence of the statistical approach inherent in medical research and practice, and it was inevitable that statistical methodology should have become as it is today the very foundation on which scientific inquiries in Medicine stand. Information on individual persons is the basis of all inquiries and this is furnished by field, laboratory, clinical and pathological techniques of investigation. Such information, in the hands of experts and specialists in the respective fields, gives a lead on an ever-increasing variety of factors to be reckoned with and a clue to the nature of their

* A talk given to the Alumni Association of the King Edward VII College of Medicine, Singapore and the Faculty of Medicine, University of Malaya on 27th October, 1959.

influence on the individual's state of health. Then statistical methods are applied, have to be applied, to follow up on these clues, to confirm or reject one's pet theories, beliefs or impressions, and to render them acceptable as contributions to the evergrowing body of medical knowledge through the medium of scientific data. Except for clinical-pathological case reports, scientific papers in the medical journals today reflect the use of statistical methods of one nature or another.

But while scientific research in Medicine has come to depend so much on statistical methodology, there is still very widespread confusion, vagueness and misunderstanding about this subject, even among many who use it.

THE PHYSICIAN'S INITIAL IMPRESSIONS OF STATISTICS, OR THE ORIGIN OF THE 'TWO-STAGE' ATTITUDE

The undergraduate medical curriculum cannot be expected to contain enough formal instruction on the subject to bring home the full implications of its scope and importance. The findings contributed by past research are taught as part of the respective subjects to which these findings belong, and even if the details of the research methods involved are explained, the medical student does not recognise their statistical nature as such simply because the lecture is ostensibly on heart disease, or neoplastic disease, or bacterial or viral diseases, and so on. He does not realize, for example, that the dividing line between normality and abnormality is determined by statistical principles; that differential diagnosis is a statistical process involving probabilities; that it is possible to recommend a treatment of choice in a particular situation only because statistical investigations have shown that this particular treatment has the best chance of success in this situation or that it has the least chance of producing certain undesirable side effects.

By the psychological process of association of ideas, the student's understanding of the scope of statistics becomes limited to the contents of only those lectures that specifically bear the label Statistics. Thus, he gets the impression that statistics is only concerned with the computation of this and that death rate or some other rate; with the uninterestingly mathematical ramifications which go by some such name as Standardization or the Life Table; with the mechanical juggling of a whole lot of incomprehensible records into tables and diagrams; or with splitting hairs over the wording of a definition. All of these topics are seemingly so coldly detached from the focus of his interest in the fourth and fifth years—the hospital ward—that one must excuse him for decrying statistics as being a "dry" subject hardly connected with the practice of medicine. The fact that such "statistics" lectures are given by a non-medically qualified person does not help matters very much.

By the same process of association of ideas, this limited impression is gradually supplemented by subsequent encounters with the word "statistics" in various other contexts, such as medical papers, books, conversation and, by no means the least important, the daily newspapers.

What impressions of statistics does he get from these sources? Books on public health frequently contain a chapter on "Vital Statistics", the contents of which leave him believing little different from what he did before, namely, that statistics consists of the definition and explanation of rates and more rates, the arithmetic of calculating the mean and median, the Life Table, and standardization. Books on medicine and surgery rarely use the word 'statistics' strengthening the impression that it has no place there. Books on statistics with few exceptions abound with figures, tables, diagrams, mysterious algebraic symbols, formulae, equations and more tables at the back, interspersed throughout by a peculiarly technical language which is difficult to comprehend. This strengthens the idea that statistics equals mathematics and, if that is not enough, there is further confirmation in the section on vital statistics in the Ministry of Health's Annual Report which gives more tables and diagrams and rates.

Next, let us turn to scientific medical papers. Here the word statistics appears usually in two forms, "statistical analysis" and "statistically significant". "Statistical analysis" appears to consist of doing something to the data, but what exactly is done is not clear because it is not explained in the paper itself. The meaning of "statistically significant" is likewise not clear, but because it seems to be a prerequisite for some finding or other to be "statistically significant" before it can be taken as proven, the reader develops a great deal of respect for this phrase.

In newspapers, the word 'statistics' nearly always refers to figures extracted from official reports. I shall not discuss whether or not such figures really show what the newspapers say they show, else I might be sued for slander. But suffice it to say that newspapers give the strongest impression to the largest number of people that 'statistics' only means 'figures' — not excluding the figurative meaning of the words in the references to external features of the human anatomy.

The total effect of all these impressions is the confusion and misunderstanding to which I alluded earlier. On the one hand, 'statistics' seem to be unreliable—some people actually believe that statistics tell lies and can be made to prove anything—so how can statistics be the medium of scientific inquiries. On the other hand statistical analysis and statistical significance seem to be necessary evils of a mathematical nature to which data must be subjected in order to produce conclusions respected by the scientific medical world. In the effort to reconcile these two aspects, one tends to think that statistical analysis is some sort of a cure for inherently unhealthy figures. This is a completely wrong approach and is particularly dangerous because it tends to divide the process of scientific inquiry into two separate and independent stages: (1) the amassing of data (which is not apparently a statistical problem), and (2) the analysis of these data (which is apparently the statistical problem). I refer to this, for short, as the 'two-stage' attitude.

EFFECTS OF THE 'TWO-STAGE' ATTITUDE

This 'two-stage' attitude to statistics manifests itself in a variety of undesirable ways. For example, a person may attempt to evaluate a particular drug by ploughing through records of patients who had been given it, believing that when he has accumulated sufficient relevant data he can then start thinking about how these data can or should be analysed to yield some conclusion on the value of the drug. Another example might be this: an investigator, doing a so-called 'survey', amasses a formidable volume of data and then brings it to a statistician with the question "What do I do with them now?". I had an interesting encounter with a person who had actually completed both stages himself when he came to me and asked "Can you please explain to me what I have done?"! But the most undesirable manifestation is shown by a person who, having found that his data do not support his intended arguments, brings them to a statistician and asks him "What's wrong with these figures?".

These approaches are to be expected in view of the comparative absence of basic training in the scientific approach to inquiries at the undergraduate level. The pity is that in the examples just mentioned, statistical advice is sought too late to be of more than very limited help. In such cases, the chances are that his inquiry is doomed to failure or, at best, to produce some result or conclusion that can be shattered by criticisms of the basic data themselves, however sound may have been the methods of statistical analysis that were applied at stage 2.

Scientific inquiries based on unsound statistical principles in the initial stages of the collection of data are unproductive and result in a tremendous waste of time (by committing months and even years to a project); waste of effort (by committing a sizeable body of personnel on it); and, of course, waste of money (through equipment and personnel). It certainly is also not conducive to improving the physician's frame of mind in respect of statistics and statisticians. As the Editor of a much respected medical publication once put it, "it is exasperating, when we have studied a problem by methods that we have spent laborious years in mastering, to find our conclusions questioned, and perhaps refuted, by someone who could not have made the observations himself. It requires more equanimity than most of us possess to acknowledge that the fault is in ourselves". Not so long ago in the course of discussions on the planning of a certain survey, someone was heard to say: "Whatever we do, statisticians will knock holes in it anyway". I wonder how many physicians have not thought so at one time or another.

Happily, not all of us suffer very long from the ill-effects of the 'two-stage' approach, but it will be appropriate to remind ourselves (particularly as we are increasingly prone to using methods of analysis given in statistics books rather like using recipes given in cookery books) that the limitations to our conclusions are imposed by the quality of our basic data themselves, and not by the subsequent analysis. Fisher has summed up the situation in this statement taken from his book 'Design of Experiment':—"Modern statisticians are familiar with the notions that any finite body of data contains only a limited amount of information, on any point under examination; that this limit is set by the nature of the data themselves, and cannot be increased by any amount of ingenuity expended in their statistical examination: that the statistician's task, in fact, is limited to the extraction of the whole of the available information on any particular issue. If the results of an experiment, as obtained, are in fact irregular, this evidently detracts from their value; and the statistician is not elucidating but falsifying the facts, who rearranges them so as to give an artificial appearance of regularity."

THE PROPER PLACE OF 'STATISTICS' IN SCIENTIFIC INQUIRY

Statistics therefore cannot be, and are not, confined only to the analysis of data independently of the collection of data. These are but two of the several aspects of the scientific dicipline implied in the word "statistics", these several aspects being intimately bound together by one overriding consideration—the aim of the scientific inquiry. To be successful in terms of fulfilling its specified aim or objective, any piece of scientific inquiry in medical research demands that the statistical problems involved be recognized and solved before the inquiry is physically embarked upon. For it is the very design of the inquiry on a sound statistical basis which renders the resulting data capable of being analysed in a particular way to achieve the intended objective of the inquiry.

I mentioned at the beginning the multiplicity of factors involved in the production of individual variations. In order to bring the relevant factors under control in a scientific inquiry, those that require to be measured must be subjected to rigid definitions and criteria of measurement. Those that are not to be measured, and in particular those that are as yet unidentified, must be controlled by appropriate sampling techniques. For example, in the case of sample surveys, the sampling technique must be such that the relevant factors in the population concerned have the best chance of being represented in their respective proportions and combinations in the sample taken from that population. Otherwise the findings yielded by the sample are not necessarily valid for the population. Or again, in comparative studies such as therapeutic or prophylactic trials, the two comparison groups must be so set up as to have the best chance of being identical in respect of all relevant factors collectively, except for the treatment or the vaccine, whichever is concerned. In this connection we might note the role of the placebo in 'blind' trials. The very act of receiving a pill or an injection and the physician's knowledge that one has received it and another not are important factors liable to affect the patient's response as well as its evaluation. These factors must be equalized in the treatment and the control groups firstly, by giving a placebo to the control group which is identical in all ways except the absence of the therapeutic agent which we are evaluating, and secondly, by concealing from the physicians who are involved in assessing the response the nature (genuine or placebo) of the treatment given to any particular person included in the trial.

The statistical procedure which ensures that the factors over which we have no control have the best chance of being equal in two groups of persons is referred to as 'randomization'. This word has a rather special meaning in the statistical context. I sometimes wonder whether it would not be more advisable to invent new words to describe certain statistical concepts, than to use common words in a specialized technical way which conflicts with their common-place meanings. Now if you said that I talked randomly, you would presumably mean that I talked haphazardly, heedlessly, carelessly, and without aim, purpose or any evident order of thought. Similarly, we tend to think that choosing people randomly for investigation means taking persons from here and there, haphazardly, without any ordered choice. This is not the meaning in which the word is used in the statistical context, in which a random sample means a sample chosen by a strictly diciplined procedure which ensures that every single member of the population (from which the sample was drawn) has had an equal chance of being included in it.

The statistical advantages accruing from random sampling in this context is far reaching. But, briefly, we may take note that random sampling is the only means of ensuring that chance variations between groups can be predicted and calculated precisely and expressed in terms of statistical probabilities, which in turn enables findings from the sample or samples to be evaluated objectively. We may also note that even very large numbers of haphazard observations will not produce conclusive results in the way that much smaller random samples will. Let me give an example of the limitations of non-random samples in the context of tuberculosis surveys. The prevalence rate of tuberculosis in a group, even as large as 50,000 persons who voluntarily attend at mass X-ray Diagnostic Centres, is not necessarily valid for the population as a whole. This is because there may be social, psychological, occupational and other factors which may make certain people fear the discovery of tuberculosis in them, and this group, who do not volunteer, may have a much higher prevalence of disease than the group who do volunteer for examination. That is to say, those who are examined are not comparable with those who are not, because of a bias in the factors involved. Therefore, even though the sample is as large as 50,000, the prevalence rate obtained is not a valid estimate for the whole population. It is also important to note that we cannot even say how far off this observed prevalence rate is likely to be from that of the whole population, simply because we do not know the extent to which the bias in the factors operates in this particular situation. On the other hand, if we take a random sample of say just one-tenth the previous number, that is 5,000 persons, track down and examine every single one of them and determine the prevalence rate for this random sample, this rate is a valid basis for generalization to the whole population; and furthermore the margin of error of our estimate based on this rate can be precisely determined, because this time it is dependent on the size of sample alone and not on some unknown bias in a factor or factors. Thus the random sample of 5,000 produces conclusive results where the non-random sample of 50,000 volunteers does not.

It is true that the mechanism of designing a random sample and of tracking down the chosen persons is administratively more complicated than simply setting up an X-ray Diagnostic Centre and asking volunteers to turn up of their own accord. This is probably what deters investigators, whatever their field of study, from seeking statistical assistance at the planning stage of an inquiry. I have on more than one occasion heard this comment: "If we consult a statistician on the method of inquiry, he will only make an otherwise simple inquiry unnecessarily complicated". The object of controversy in that remark is the word "unnecessarily". Is the additional "complication" at the initial stage "unnecessary" if it means the difference between fulfilling and not fulfilling the objective of the inquiry; between producing conclusive and inconclusive findings; between spending less money more wisely on 5,000 examinations and spending more on 50,000 with the concomitantly greater difficulties of following-up and recording, processing and analysing data on 45,000 more persons, all to little avail in the context of a scientific inquiry.

There are, of course, in addition to this question of 'who to study', many other things which must be carefully examined and provided for in order to ensure that the inquiry will be carried out on a sound scientific basis. Some of the problems involved are more clearly of a statistical nature, while there are others which, by virtue of the fact that they concern specialized laboratory or clinical techniques, do not appear to be the concern of statisticians. But in so far as this latter group of specialized techniques are instrumental to the production of data to be used in the study of a particular problem, they present as much of a statistical problem, if not more, than the first group.

Specialization, which is so necessary for the study, practice and advancement of each field of technology, carries with it the potential danger of each specialist not recognizing the problems in other fields outside their immediate interest of focus of attention. This is probably true in the general sense, but I am particularly referring to the different aspects of scientific data involved in any inquiry. If the inquiry is being carried out by a bacteriologist, he tends to be preoccupied with the bacteriological techniques of obtaining specimens and of isolating and identifying the bacteria, or with the techniques of cultures and complement fixation tests and so on. Similarly, a chest physician tends to be preoccupied with the techniques of diagnosing and treating the cases of, say, tuberculosis—so also, the other clinical and surgical specialists, in their respective fields. Even in their own fields, how often do they seriously examine and evaluate the chances of error in each test, each diagnosis, each reading or each classification. The evaluation of such errors arising in relation to the special techniques employed is fundamental to the evaluation of the resulting data for whatever purpose. If the specialist undertaking the inquiry takes positive steps to minimize these errors by ensuring that everybody concerned adheres to rigidly specified and objective criteria, then the inquiry will be placed on a scientific basis. It is a necessary duty of a participating statistician to see that this is done so that the resulting data will not be open to criticisms of unreliability, of irreproducability or of the presence of unknown 'observer differences' arising from subjective factors, lack of standardization of techniques and criteria of evaluation. A glaring example of a faux pas in this direction was seen in a particular tuberculosis survey at the end of which the prevalence rate of 'active' cases was compared between two geographical areas in each of which a different chest physician handled the cases, and the number of 'active' cases was taken to include 'all cases which were recommended for treatment by the chest physicians'. A considerable difference was found between the prevalence rates of the two areas. To what extent was this difference due to the two physicians taking different attitudes towards 'when they should recommend a particular person for treatment as a case of tuberculosis'? So long as the criteria for doing so remains unspecified and can include variable subjective factors, we cannot, from a comparison of the resulting prevalence rates, say with any degree of certainty what the 'real' difference is between the prevalence of tuberculosis in these two areas. Note that we cannot even say what the prevalence in any one of the areas is in this example.

There are other items of information which must be included in any inquiry. If we intend to carry out analysis by age and ethnic group, we have to record the age and ethnic group of each person concerned. If we simply tell our clerk or our nurse to 'keep a record of each person's age and race', we are again going to land in trouble later on in the interpretation of the resulting data. We will be faced with such questions as "how many of the people recorded as being 20 years of age are in fact 20 years of age by the conventional meaning of age in terms of completed years of life or, as it is commonly called, by the 'English' reckoning?" In order to forestall such queries and the resulting uncertainties about the reliability of data, we have to define all our items of information beforehand — not at the end of the inquiry when we are writing the report—and issue explicit instructions not only on their meaning, but also on exactly how they are to be asked for or calculated, as well as on the approximations and classifications which are to be used. I have in mind a particular inquiry involving the recording of weight in which some records showed weights to the nearest pound, some to the nearest half-pound, some to the nearest quarter-pound, some to the nearest ounce and so on. They present innumerable problems, all of which could have been ridiculously easily prevented at the very beginning by the issuing of simple, written, explicit instructions.

Seen in this light, statistics therefore has a tremendously important 'preventive' role in all scientific inquiries in addition to providing the appropriate methods of recording, processing, analysis, presentation, and interpretation of data. Many of these methods are applied after an inquiry is completed, but their problems must be recognized and resolved before the inquiry begins if it is to be put on a sound scientific footing. I have not been able, in this short space of time, to attempt to give more than the merest suggestion of the nature of one or two of these problems; but if I have at all helped to create a broader outlook to statistics among some unsuspecting victims of the 'two-stage' attitude, I shall have made a useful contribution.