SOME PRACTICAL ASPECTS OF MEDICAL STATISTICS

By R. H. GORRILL, M.D.

Reader in Bacteriology, Guy's Hospital Medical School, London, S.E.1

The use of mathematical methods in analysing and reporting medical data is increasing. This is particularly seen in certain types of clinical trial where the whole experiment is designed from the start to yield results than can be easily tested by statistical methods. The purpose of this article is to discuss some statistical methods in common use without examining the detailed mathematical argument behind them.

The outstanding feature of most medical experience is the variability of results. This is seen most obviously in the clinical field where individual cases often respond to a disease in unexpected ways. It is by seeing a large number of cases that the doctor builds up in his own mind a picture of the average behaviour which he comes to expect. With increasing experience he gets to know not only the average but also the extremes. This allows him to recognize the genuinely unusual case.

The statistician attempts as far as possible to arrive at a measure of this natural variation so that he can give in numerical terms what is in fact the underlying knowledge of the experienced clinician. In order to arrive at a measure of the frequency of given events a considerable number of cases have to be examined and recorded so that a body of evidence can be accumulated.

One of the important features of this sort of work is accurate recording and then devising methods of reporting the results in as compact a form as possible. The best way of doing this is by tabulating in suitable class intervals so that the information is not too condensed. With any mass of data about 20 class intervals are usually necessary. There is in addition the use of block diagrams, graphs and in some cases scatter diagrams. While these convey the results very easily to the reader they often fail to give sufficiently precise information to anyone working in the same field. For this reason tables are the most favoured by statisticians. It should be realized that there is often great difficulty in medicine in arriving at even the basic figures, an obvious example of this

being the duration of malignant disease at first diagnosis. Here patients vary in their awareness of the presence of the neoplasm and often give histories which are quite obviously at variance with the size of the tumour found.

Having collected the information one now has to put it into as concise a form as possible. The obvious first step is to arrive at the average. The average however is not entirely satisfactory because while in some cases it conveys all we need to know, in many others the variability is so important that the average expressed as such is almost valueless.

In order to measure the variability we want some simple value which is both universally agreed and simple to use. The most common way to measure the degree of variability is to calculate the Standard Deviation. This is done by taking each value in turn and finding out how much it differs from the average (or mean). This difference is then squared and all the squared deviations are divided by the number in the group and the average squared deviation is found. This is known as the variance. The square root of the variance is known as the Standard Deviation. With small samples it is usual to divide the sums of the squared deviations by one less than the number in the sample. The effect of taking the square of the difference is to very greatly weight the big differences.

At this point two asides may be allowed. (i) The problem of degrees of freedom which is involved above. In small groups only one less than the number in the sample is unknown because once all but one of the readings are known the last is fixed by the total. This means that the degrees of freedom are one less than the sample size. Degrees of freedom are involved in many statistical procedures and are a definite problem to many students. In general it may be said that the number of degrees of freedom is either quite clear or else well known because of previous experience with the method. It is therefore only in a few cases an important obstacle. In these it

is usually possible to find a similar worked example in a text book or else to go to a trained statistician. One can do a considerable amount of simple statistics well without properly understanding the concept of degrees of freedom. (ii) The variance described above is a measure of the variability of the data. Now many factors contribute to this variation and in some cases it is possible to split up the variance and decide how much each factor contributes to it. By this method, known as analysis of variance many more advanced statistical problems may be solved.

The value of the calculation of the mean and the standard deviation of a mass of data is seen if one considers the following idea. It is quite possible to write mathematical formulae which describe not only straight lines, circles and other simple shapes but also most complex forms. With such a formula it would be quite easy for example to calculate the area of a circle lying between two parallel lines.

Now if we plot certain types of biological data, taking the frequency of the observed value against the value, we get what is called the normal curve. This hump-backed curve is too well known to need reproduction here, but it does describe the distribution of a very large number of biological values.

Now such a curve can be described by means of a mathematical formula and as in the circle analogy it is possible to calculate the area enclosed by two parallel lines cutting the curve. It is found that the area lying on either side of the mean enclosed by one or more standard deviations is the same for all normal curves. The values are, 68 per cent. of the area (and therefore of the observations) lies within ± 1 SD of the mean; 95.5 per cent. lies within ± 2 SD and 99.73 per cent. within ± 3 SD. Looking at this from the other point of view, a value greater or less than the mean ± 1 SD (i.e., 100 per cent.-68 per cent.) will occur in 32 per cent. of observations, greater or less than the mean ± 2 SD (100 per cent.-95.5 per cent.) in 4.5 per cent. or 1 in 20 observations and greater or less than the mean ± 3 SD (100 per cent.-99.73 per cent.) in 0.27 per cent. or 1 in 370 observations. Hence by calculating the SD and the mean we arm ourselves with a valuable statistical measure.

At this point it is as well to remind readers that all that can be calculated is the probability of the result being observed by chance. Hence if we were to make an observation say more than $3 \times$ SD above the average we would be very strongly inclined to think that as this only occurs by chance once in a thousand times it was significant. It should be noted that the SD is used to measure the area around the mean, e.g., $\pm$ SD includes 99.73 per cent. of the area; hence .13 per cent. of it lies above and .13 per cent. below, i.e., about 1 value

in 1,000 is larger and 1 value in 1,000 is smaller by chance alone.

The concept of the normal curve can be extended in several ways helpful to medical workers. If we take a number of samples of the same variable and calculate the mean of each sample and its corresponding SD we shall accumulate a series of estimates of what we term the true mean and the true SD. By this we imply that the variable measured has an absolute mean and SD and each sample gives a measure of it. If we plot all the sample means as we did the individual values above we find the sample means are normally distributed with a SD given by the true SD divided by the square root of the sample size. This SD of the mean is known by the term Standard Error of the mean. Since we do not know the true SD of the whole population we take the SD of the sample as the nearest approximation. Consider a measurement in a sample of 25 giving a mean of 50 and a S.D. of 10. The S.E. of the mean is $10/\sqrt{25} = 2$. Here we would expect that a range of 50 ± 6 i.e., mean ± 3 S.E.) would include 99.73 per cent. of means of samples of 25. The similar figures for single observations are 50 ± 30 (i.e., mean ± 3 S.D.). This shows in mathematical terms that samples give much less variation than single values. As well as considering the relationship between the sample mean and the true mean we can also compare two samples directly. Here the pooled S.E. of the two is used to measure the significance of the difference between the two means.

Because of the importance of the normal distribution it is used wherever the data even approximate to the normal form and often with large samples the distribution becomes more nearly normal. This is well seen in the case of proportions. Often in medicine we know that in a given disease say 20 per cent. die, hence $p = .2$ (the probability of death) and $q = .8$ (the probability of survival). If we take two cases, we find that the probabilities of both dying, one dying, one living and both living are given by $(p + q)^2 = p^2 + 2pq + q^2 = .04$ (probability of both dying 4 per cent.) + .32 (probability of one dying, one living 32 per cent.) + .64 (probability of both living 64 per cent.). Now this argument can be extended to any number of cases (n) by the formula $(p + q)^n$. Rapidly the arithmetic becomes impossible but with large values of n (say over 20) the distribution becomes so nearly normal that we find that the mean is p (i.e., 20 per cent.)

and the S.E. of the mean is $\sqrt{\frac{pq}{n}}$ which in 20 cases gives a S.E. of 9 per cent. This would mean that in a series of 20 cases all might survive about once in 50 trials by chance alone. Hence it might be

very difficult with only 20 cases to show the value of a new drug.

In a number of cases we are not confronted by a series of measurements which we can arrange in an orderly array and fit to a normal curve but a series of frequencies; either where the frequency is too low for easy use of the p.q. system or where the frequencies of a number of characters are recorded in two or more groups.

These problems are met by use of the Chi squared test (X^2). The value of X^2 which is the sum of the deviations from the expected values squared and then divided by the expected number is not of course distributed in the same way as the S.D. of the normal curve. The values of X^2 for different degrees of freedom at different levels of probability have been calculated and are available as a table. X^2 is a very simple statistical process and is one of the most frequently used methods in medicine.

As an example we may consider the results of anti-typhoid inoculation, 6,815 people were immunized and of these 56 developed typhoid fever during the period of observation: 11,668 people were similarly exposed and not being immunized served as controls, among them 272 cases were reported. These figures are set out in the table.

	Attacked	Not Attacked	Total
Inoculated . . .	56	6,759	6,815
Non-inoculated . .	272	11,396	11,668
	328	18,155	18,483

Now the only problem is the calculation of the expected number. This is done as follows: in a total population of 18,483, 328 cases of typhoid were seen. If immunization was without effect we would expect the 6,815 people in this group to have their share of the cases, i.e., $328 \times \frac{6,815}{18,483}$
 $= 120.93$. Hence in the first cell the observed value is 56, the expected is 120.93

$$= \frac{-64.93^2}{120.93}$$
 By a similar argument the value for

each cell is calculated and added together giving in this example a total of 56.2. In this case there is one degree of freedom and the value of X^2 greatly exceeds that given for a probability of 1 in a 1,000 (it is in fact about 1 in 2,000,000). In view of the very low probability we think it is

extremely likely that immunization is responsible for a decrease in the incidence of typhoid fever.

One other problem that can be briefly considered is that of correlation or the relation of one variable to another. In its simplest form it can be studied by the use of a scatter diagram, the value of one variable being marked along the X axis and of the other along the Y axis. This can be extended to the negative axes if necessary. Each patient will give a value for both X and Y and this will be shown by the corresponding point on the diagram. When all the pairs of values have been entered the points may arrange themselves in a fairly narrow belt diagonally across the paper, suggesting that increase of one is accompanied by an increase (or decrease) of the other. The scatter diagram gives one a visual picture of the correlation. If none appears to the eye in general none is shown by calculation.

Now by eye it is possible to draw a line through the dots giving the general trend. It is possible to calculate what we call the best fitting straight line instead of drawing it by eye. This is done by calculating the line which has the smallest sum of squares of differences, i.e., we add up the square of the distance of each point from our line. When that sum is smallest the line is the best fitting line. The effect of the sums of squares method is to give more weight to the big differences, hence the best fitting line drawn on the graph is not always the same as the one we have fitted by eye. As well as drawing the best fitting lines it is also possible to calculate the dispersal of the data about the line and so to arrive at the correlation coefficient. The method is arranged that perfect correlation gives a value of ± 1 and no correlation a value of zero. Naturally, neither of these values is ever seen and would in fact be highly suspect if it were, as biological data are too variable to give either result. When we have calculated the correlation coefficient the next step is to see whether the value is significant or not. This is done by calculating the S.E. of the mean which is given by $1/\sqrt{n-1}$, i.e., with 26 pairs of readings the S.E. is .2. The argument is then as follows, if there were no correlation then the true value is zero. Owing to the variability of data there will be some scatter but 68 per cent. of the values will lie within 1 S.E. (i.e., $\pm .2$) of zero and 95.5 per cent. within 2 S.E. (i.e., $\pm .4$) of zero. Hence a correlation coefficient of $+.4$ in a sample of 26 pairs would occur by chance only once in 50 times.

It should be pointed out that correlation methods are often criticized when the presence of correlation is taken to show a relationship. If two values are correlated they may be related or on the other hand they may be quite unrelated but moving in the same direction for two quite different

reasons. A crude example of this is the correlation between steel production and divorce. Both happen to have risen in the last 50 years. It is unlikely that they are related.

When one turns to consider the problem of material published in the form of rates the difficulties are fairly well known. In general crude rates are of little value for comparison because they are not based on identical populations. The population may differ in two main ways; firstly the severity of the cases may be different, secondly the age and sex structure of the population may not be the same. These difficulties can be overcome in various ways, in the first case an agreed classification of cases can be drawn up and rates per class reported. This is often used in some types of malignant disease. It is unfortunately very difficult to get really rigid classes and as a result doubt is often expressed as to the value of comparisons of rates between different centres particularly if they are in different countries.

If the cases are not difficult to classify then the problem of age and sex structure can be overcome by the use of standardized rates. Here the population under examination is broken up into age classes, say at five-year intervals, and into male and female. The rates for each class are then calculated and from these rates the number of cases in an agreed standard population is discovered. Such a population can either be constructed artificially or one of the early census returns used. From these figures the overall rate in this model population may be found. This is the standardized rate. There are a number of similar methods that are worth study.

All these methods are devoted to the study of material already available. The tendency today is to look ahead and to plan the trial or experiment so that it will yield results that lend themselves to analysis. Furthermore the analysis is usually much more easily carried out than in the case of unplanned data.

Before starting a clinical trial or to design an experiment it is desirable to put down exactly what one wishes to find out. Having decided this the next problem is to decide the manner in which the results will be obtained. To take a simple case, one has a drug believed to relieve headache, here the problem is to measure its efficiency and one would do this by giving the drug to patients with headache and measuring the degree of relief. At once one is faced by the problem, what sort of headaches will you treat and how does one score the result? It is this previous consideration of the difficulties which is so important in trial design.

Only too often it is difficult or impossible to

develop a method of scoring the severity of the disease and so the response to treatment. This is particularly true of chronic conditions where cure is unlikely and where there is considerable fluctuation, often with the seasons of the year. This aspect of clinical trials is a challenge to the medical worker and is as important as the final analysis. When the scoring method has been decided there are roughly two ways of using controls. In many conditions other cases have to be compared but in some trials the patient himself can act as the control by giving alternate periods of drug and placebo. When control cases are employed it is usual to allocate them by some random system rather than to use alternate cases. In general if the clinician can be kept unaware which cases are controls so much the better.

When cases are accepted for a trial it is usual to record a number of apparently unrelated facts, the purpose of these is merely to show that treated and control groups are as near identical as possible. For example it is usual in a trial of a vaccine to show that the two groups had been immunized against say diphtheria with the same frequency. If there was a marked discrepancy as shown by a χ^2 test one would wonder if the cases had perhaps been allowed an element of self selection.

This paper is not a suitable place to discuss the ethics of medical trials except to say that one must be most careful to explain to patients or their guardians that they are taking part in a trial and to get their free consent. Failure to do this may lead to hostile criticism. A further point that must be faced is that if the results of the trial are definite rather earlier than one expected then the trial should be drawn to a close. Occasionally one hears of trials that had gone on longer than was necessary and in the case of serious infections this could place the control group in unnecessary hazard.

In general the only remaining problem facing the amateur trial designer is to decide how many cases to include in the trial. There are methods given in the text books for calculating the groups size if the level of probability is decided and the magnitude of the expected difference is known. Of course the smaller the effect the larger the group required to show it. Finally if the difference is very small there may not be enough cases available to show it up. Between these limits one has to make a rough rule and in general groups of the order of 40 to 50 will show up most important differences.

Once the trial is finished it is usual to record the data for each case on a separate card. This gives a robust system which allows for the many different sortings that will be made. Cards about 5 in. by 3 in. are usually adequate. For trials

of over 200 cases (i.e., 200 cards) many workers prefer to use a system where the data are coded into punch holes at the side. Then sorting is done with the aid of a knitting needle. Finally large trials make use of the punched card which is then machine sorted.

This short paper is written in the hope of encouraging a general interest in statistical methods. There are many excellent books and almost all workers will find one that they can

understand. It is worth taking some pains to find the most suitable book as it can make a great difference to the speed with which one picks up the methods. In general books that contain exercises with the worked answers are the most useful as they allow a ready test of the degree of understanding achieved. However well one may fare with the details of statistical procedure the most important thing is to develop a statistic outlook.

ANAESTHETICS

(Special number of *Postgraduate Medical Journal*, September, 1955)

Price: 3s. 10d. post free

THE MODERN VIEW OF ANAESTHESIA

G. S. W. Organe, M.D., D.A., F.F.A.R.C.S.

THE PRODUCTION OF UNCONSCIOUSNESS

B. G. B. Lucas, D.A., F.F.A.R.C.S.

ANALGESIA

J. B. Wyman, M.B.E., D.A., F.F.A.R.C.S.

MUSCLE RELAXATION IN SURGERY

Angus Smith, F.F.A.R.C.S.

CONTROL OF THE BLOOD PRESSURE AND CONTROLLED HYPOTENSION

C. F. Scurr, M.V.O., D.A., F.F.A.R.C.S.

CHLORPROMAZINE AND ALLIED SUBSTANCES

John Beard, M.D., D.A., F.F.A.R.C.S., D.C.H.

CONTROLLED HYPOTHERMIA

E. J. Delorme, M.D., F.R.C.S.(C.)

MANAGEMENT OF THE APNOEIC PATIENT

Ronald Woolmer, D.A., F.F.A.R.C.S.

THE USE OF ANTIDOTES IN ANAESTHESIA

B. A. Sellick, D.A., F.F.A.R.C.S.

Published by

THE FELLOWSHIP OF POSTGRADUATE MEDICINE

60, Portland Place, London, W.1

RUTHIN CASTLE, NORTH WALES

A Clinic for the diagnosis and treatment of Internal Diseases (except Mental or Infectious Diseases). The Clinic is provided with a staff of doctors, technicians and nurses.

The surroundings are beautiful. The climate is mild. There is central heating throughout. The annual rainfall is 30.5 inches, that is, less than the average for England.

The Fees are inclusive and vary according to the room occupied.

For particulars apply to THE SECRETARY, Ruthin Castle, North Wales.

Telegrams: Castle, Ruthin.

Telephone: Ruthin 66