Chance and the blood count

Donald Mainland

By the early 1930s the science of statistics as we now know it was rapidly developing. Correlation-regression and analysis of variance were well known to British and North American statisticians. Nevertheless, as this article explains, statistical analysis was rarely applied to clinical results. Donald Mainland was a British statistician who spent some time in Canada. His plea for judicious use of statistical analysis is probably the first of its kind in North American clinical literature, yet it is as timely today as it was 60 years ago.

Suppose we have a patient suspected of acute appendicitis, whose differential count of 100 cells shows 70 per cent of polymorphonuclears, and an hour later, 80 per cent. Is this a significant change, or may it be due entirely to chance variation?" These sentences are quoted from a recent article by Barnett[1] on "The Unavoidable Error in the Differential Count of the Leukocytes of the Blood". It is a curious fact that, at this stage in the history of blood counting, an article of that kind should still be required in order to explain to laboratory workers and clinicians how to make allowance for the "chance" errors in the differential count.

If two differential counts are made on the blood of the same person at different times everyone recognizes that the difference between the counts may be due to: (*a*) actual differences, physiological or pathological, in the person's blood; (*b*) differences in the observer or his technique. It does not seem, however, to be so clearly realized that a difference between the two counts should be expected even if the blood had exactly the same proportions of cells on the two occasions, and even if the technique and the observer's reactions were perfect.

1. *Barnett, C.W.*, J. Clin. Invest., *1933,* ***12:*** *77.*

This article was published in the June 1934 (vol. 30, no. 6) issue of CMAJ, *pages 656 to 658.*

An argument by analogy may elucidate the point. Let it be supposed that many thousands of coins are tossed on to the floor and then picked up two at a time, purely at random, that is, without being previously looked at and so as not to give special preference either to those with head uppermost or to those with tail uppermost. If there is on the floor an exactly equal number with head up and with tail up, an average sample of two coins will show this also, but no one will be surprised if some of the samples show two heads or two tails. The differences are unavoidable, and are called "variations due to random sampling". Exactly the same kind of variation is

Statistical methods . . . are completely ignored by many clinicians.

to be expected in differential blood counting, even if the actual proportion of cells remains constant and if the technique is perfect. Hence, before taking account of other factors, it is surely desirable to know how far the differences between counts can be attributed to variations in random sampling. The methods that give the answer to this question are statistical methods, and Barnett's paper illustrates the use of certain of these methods in the differential blood count. Now all medical science, carried on by the bedside or in the laboratory, is based on sampling of one kind or another, and where samples of blood, urine, bacteria, drugs, animals or anything else are taken, random sampling variations are met in addition to any other variations that may be present. It is all the more remarkable, therefore, that statistical methods are still looked at askance by a number of medical scientists and are completely ignored by many clinicians. Anyone who cares to apply elementary statistical tests to medical literature

← For prescribing information see page 324

finds himself agreeing in substance with the verdict of Dunn[2], who, in 1929, reported that out of 200 quantitative medical-physiological papers in current American periodicals over 90 per cent required statistical methods and did not use them, and that in almost 40 per cent conclusions were drawn which could not have been proved without setting up adequate statistical control. Medical research does not in this respect compare very favourably with agricultural and industrial research.

Our most exact knowledge of pure random sampling is based on our experience of coin-tossing and dice-throwing. In the coin-tossing experiment just referred to, whenever we take up a coin at random we have an equal chance of taking one with the head up and one with the tail up. If we take them in samples of two at a time we shall expect the following possibilities: —

Head–Head	Head–Tail
Tail–Head	Tail–Tail

This shows that the average will contain a head and a tail, while two heads will occur in about one-quarter of the samples, or, in other words, the chance of finding such a sample is one in four. As the numbers of the samples become larger, this method of writing out the possibilities becomes increasingly tedious and mathematical methods are employed. It must be noted, however, that these methods do not involve any further hypothesis regarding the phenomena themselves. The mathematical methods are merely tools. By means of them, tables are prepared which are applicable to the larger samples, and they can be applied to blood sampling as well as to any other sampling, for it must be remembered that the statistician does not claim to teach the medical man about blood cells as blood cells, but as objects liable to random sampling variation.

Let it be supposed, for example, that the polymorphonuclears and the non-polymorphonuclear leucocytes are present in the blood in equal numbers, and that a number of drops are taken and on each drop a differential count of 100 cells is made. The tables can be used to show what variations are to be expected solely as a result of random sampling, apart altogether from variation due to technique or other causes. Random sampling will account in this case for results varying between the limits 35 per cent and 65 per cent polymorphonuclears. A single line of calculation with a glance at the appropriate table is sufficient to show this. Again from random sampling variation, 5 per cent of the samples will give counts either below 40 per cent polymorphonuclears or above 60 per cent, and about one-third of the samples will give counts of below 45 per cent or above 55 per cent polymorphonuclears. Barnett's experiments, designed to eliminate all but random-sampling variations, gave results that agree with these limits.

These data and the tables referred to so far are concerned with phenomena in which the chances for and against are equal (*e.g.*, 50 polymorphonuclears to 50 non-polymorphonuclears), but investigations have shown that the same kind of calculations and the same tables can be used without serious error where the chances of meeting the two types of cells are not equal, *e.g.*, where in the total blood there are 70 polymorphonuclears and 30 non-polymorphonuclears or 10 monocytes and 90 non-monocytes.

An extension of the same principle of random sampling enables one to solve the problem quoted at the outset of this article. The most useful method developed for this purpose is called the chi-square method. This method answers the question: If the differences of 70 per cent and 80 per cent in the quoted instance were due solely to random sampling, how often would you expect to meet differences of such a size? A minute's calculation gives the chi-square value, and this, turned up in the published tables, shows that we should expect to meet such differences due to random sampling in 1 out of 10 cases. If 200 cells had been counted and the same difference found, random sampling would account for it in only 1 out of 50 cases. Such are the answers given by statistical tests. It should be clearly realized that the tests do not claim to show that random sampling was or was not responsible in any given case. They do not set out to establish causes. If the clinician or laboratory worker using the tests thinks that random sampling variation will, for his purposes, sufficiently account for the results, then he need not be concerned with other sources of variation, such as technique, which would always tend to increase the variation. But this decision is left in the hands

2. *Dunn, H.L.*, Physiol. Rev., *1929*, *9*: *275*.

of the one who is investigating. It is not dictated by the statistician.

The way to a more adequate understanding and treatment of medical data would be opened up if all records, articles, and even abstracts gave, besides averages, the numbers of observations and the variation, properly expressed, *e.g.* as standard deviation (maxima and minima being very unreliable). Still further progress would be made if some fundamental ideas were more clearly understood, namely, that the principles underlying statistical methods are relatively simple, that the commonest methods are easy to learn, that the methods can be used as an instrument without a deep knowledge of their mathematical structure, that the methods do not impart a fictitious accuracy or an artificial quality to the results, and that these methods tend very often to show that conclusions are not so definite as the unaided observer would think they were. If these things were understood, the methods would be much more commonly used, and, more important still, workers would come to recognize when they should appeal to the statistician. This in turn would hasten the coming of the day when a consultant statistician will be considered necessary in every medical centre.■

The principles underlying statistical methods are relatively simple.

Conferences

continued from page 204

Apr. 16-20, 1993: 18th Annual COACH Conference
Toronto
Steven A. Huesing, 1200-10460 Mayfield Rd., Edmonton, AB; tel (403) 489-4553, fax (403) 489-3290

Apr. 17, 1993: Education in Occupational Medicine
Hamilton, Ont.
Dr. David Muir, McMaster University Medical Centre, Rm. 3H50, 1200 Main St. W, Hamilton, ON L8N 3Z5; tel (416) 525-9140, ext. 2332, fax (416) 528-8860

Apr. 18, 1993: 5th Annual Symposium on Treatment of Headaches and Facial Pain
New York
Dr. Alexander Mauskop, Director, New York Headache Center, 301 E 66th St., New York, NY 10021; tel (212) 794-3550

Apr. 18-21, 1993: Medical Excellence in Africa — 57th Biennial Congress of the Medical Association of South Africa (MASA)
Sun City, South Africa
MASA congress convenor, PO Box 20272, Alkantrant 0005, Republic of South Africa; tel 011-27-012-329-1359, fax 011-27-012-329-1345

Apr. 23-25, 1993: Cardiovascular Nutrition — Food, Fitness and the Family: 10th Anniversary Symposium of the Sports and Cardiovascular Nutritionists (a dietetic practice group of the American Dietetic Association)
San Diego
David Berwick, Great Gatherings & Perfect Meetings, 100-34 Union St., Brighton, MA 02135; tel (617) 782-4145

Apr. 28-May 1, 1993: 9th National Conference — Invisible No Longer: Rights, Routes and Responsibilities
Richmond Hill, Ont.
Learning Disabilities Association of Ontario, 124 Merton St., 3rd Floor, Toronto, ON M4S 2Z2; tel (416) 487-4106, fax (416) 487-9489

May 1, 1993: Education in Occupational Medicine
Hamilton, Ont.
Dr. David Muir, McMaster University Medical Centre, Rm. 3H50, 1200 Main St. W, Hamilton, ON L8N 3Z5; tel (416) 525-9140, ext. 2332, fax (416) 528-8860

May 2-4, 1993: CEO Leadership Institute
Ottawa
Canadian College of Health Service Executives' Professional Services, 201-17 York St., Ottawa, ON K1N 5S7; tel (613) 235-7218, fax (613) 235-5451

May 6-7, 1993: 9th National Conference on Prescription Medicine Information and Education — Making Progress in Medicine Communication: the Outcomes Challenge
Washington
National Council on Patient Information and Education, 9th National Conference, 666 Eleventh St. NW, Ste. 810, Washington, DC 20001; fax (202) 638-0773

continued on page 242