Do Statistics Lie?

CHESTER B. STEWART, B.Sc., '38.

DURING an election campaign an old man on his way from a political meeting was heard to remark: "I am not going to vote for that candidate. I never believe a man who proves everything he says with figures."

Such an expression of doubt concerning the veracity of statistical evidence is not at all unusual, even regarding data not perhaps so likely to be garbled as the common political brand. Most medical students have at some time been reminded that statistics—like the Bible—can be made to prove almost anything, or have heard the quotation concerning "lies, damn' lies and statistics." Moreover, the critic will often quote from medical literature theories which were apparently upheld by statistical evidence, but were later disproven. Several quotations will be presented in this article and an attempt will be made to show why the writers' conclusions were unjustified.

A simple dictionary meaning of the word "statistics" is "classified facts; especially those which can be expressed in numbers, or (singular) the compilation of such facts." It is not usually the statistics which are proven incorrect, since to gain these facts the careful investigator eliminates every possible source of error. Much more frequently it is the investigator's deductions which are later questioned, and an examination of the original data will very often show that he did not correctly classify his material or did not apply tests to determine the significance of the results. Statistical data, if competently collected and suitably classified, may show very decidedly that only one conclusion is reasonable. If, however, results are less striking, other aid should be enlisted. The omission of various tests may lead the investigator to express an opinion not warranted by his results or at least leave the reader of his report with an erroneous impression.

Statistical Tests in Public Health Investigation:

The value of various tests in the analysis of Public Health statistics is very generally recognized. It can be shown, for example, that an examination of the crude death rates from cancer per 100,000 of the population does not give an adequate or correct comparison of the prevalence of this disease in the various provinces of Canada. The average rates per 100,000 per year between 1928 and 1932 show that Nova Scotia had the highest mortality rate of 113, Prince Edward Island was second with 109, and Manitoba was seventh with 85. Cancer is, however, a disease which affects most commonly the middle-aged or elderly portion of the population, and, when the above rates are corrected to take into account the age distribution of the population in each province, a more revealing comparison is obtained. Manitoba ranks first with an age specific cancer death rate of 96.1 per 100,000, Nova Scotia third with 91.7 and Prince Edward

Island ninth with 78.1. Contrary to a rather common opinion, however, statistical tests are not of use only in Public Health investigation or when large numbers are involved. In fact, they are just as important in general medical investigation and when small samples are being used. Here also adequate classification according to age, sex, etc., is often very important if incorrect deductions are to be avoided.

Statistical Tests in Medical Literature:

Within the last year the writer has examined articles in several well-known medical publications with special reference to the omission of one simple test. In a comparatively few hours of reading and calculation some ten articles were discovered in which one or more of the author's conclusions were shown by an analysis of his own data to be unjustified. At least one of these conclusions was later quoted as fact by another writer in a different publication. One wonders if many such statements are being reprinted both in articles and textbooks without an analysis of the original data to determine their basis. Then, too, one wonders about the numerous therapeutic measures which were once lauded and are now discarded. Might an analysis of the figures concerning the number of cures have shown the investigators that chance alone, without their therapeutic procedure, could have been responsible for as great a difference as that between their series and the controls?

Arterial Supply of the Brain in Sane and Insane:

In the Journal of Anatomy and Physiology, January, 1916, Dr. J. S. B. Stopford discusses the theory that criminals and insane persons are more likely than the sane to have an abnormal arterial supply to the base of the brain. He examined 118 brains from sane persons and 32 from insane, and found that 61% of the sane and 79% of the insane had anomalies of the basal vessels. These figures, he states, support the above theory. His conclusion is based apparently upon a superficial examination of the percentage of anomalies. One wonders if the results had shown 65% of the sane and 75% of the insane with anomalies, or 68% and 72% whether the same statement would have been made. In other words, must we depend upon the rather haphazard conclusion of one man as to whether or not a difference is large enough to be significant? Does it not seem absurd to take excessive pains to eliminate all sources of error from a scientific investigation, and then ignore tested scientific methods of assessing the validity of the deductions made from the results?

In medicine we must very frequently work with samples, sometimes small ones, from a general group, whether it be a blood specimen from the whole blood stream, a group of patients from the whole class with one disease, or a number of brains from a certain group of cadavera. It is very important to know that the conclusions drawn from the arrangement in this sample are accurate, but even more important to know whether we are justified in applying these conclusions to the whole class from which the sample came, or to what extent the deductions should be modified in such an application.

If conclusions are to be applicable to the whole series, the sample must be fully representative. This condition is best fulfilled if it is a random sample; that is, one chosen in such a way that chance alone determines whether or not an individual or object is included. We suppose that there must be numerous factors tending to cause anomalies of the basal vessels in foetal life. More of these factors may be active in certain groups as to race, sex, economic status of parents, etc. If a sample, such as Stopford's 118 sane, were chosen so that no particular class of individual as to race, sex, etc., was more likely to be included than any other, these factors would tend to balance each other. Such a sample would be more likely to represent the true arrangement and the conclusions drawn from it would be applicable to the whole class, provided allowance was made for chance variation. No information is given in the article to show whether or not Stopford's groups were representative, but supposing they were how should allowance be made for chance variation?

Let us suppose that a great number of brains from sane persons were examined and the results showed half with anomalous and half with so-called normal basal vessels. If we then chose at random from this series several groups of two brains, we should expect by chance alone to get some samples of two brains with normal vessels, some of two with anomalies and some mixed samples one with normal and one with anomalous vessels. The possible arrangements would be NN, AN, NA, and AA, if N and A represent normal and anomalous arterial supply, respectively. In other words, we should expect only half our samples of two brains to consist of the mixed arrangement which represents the actual condition in the whole group. There would be one chance in four of getting either two with anomalous vessels or two with normal, although each group of two brains is a random sample.

If from the same series of brains half with anomalous basal vessels and half with normal we chose groups of five, the possible arrangements would be increased. We should not be so likely to get a sample of five brains with normal vessels and none with anomalies, but we should still expect this grouping in one out of every thirty-two samples. This is simply calculated by use of the binomial theorem. Only twenty of every thirty-two samples would be expected to show either three with normal and two with abnormal vessels or vice versa, the two arrangements nearest the true ratio. If, then, an observer examined only five brains and found four with normal and one with anomalous vessels, he would not be justified in stating that this was the true arrangement, even though the group of five was a random sample from the whole series. If the true ratio were one to one, i.e., 2.5 to 2.5, the chances of getting arrangements showing as great a difference or greater than four to one would be twelve in thirty-two. Therefore, unless he is willing to be wrong in 35.7% of his conclusions his statement is not justified. Neither should he state that a group of five brains all with normal vessels does not come from a series with a one to one arrangement, unless he is willing to be wrong in 6.2% of his conclusions. If aware of this fact, he may be willing to take the risk. Anyone can use the level of accuracy he wishes to, but in biological and most statistical work no error greater than 5% is accepted. We are justified in stating that some other factor besides chance is at work only if there is less than a 5% probability that the two arrangements came from one series.

In a larger group such as Stopford's 118 sane the same conditions apply. He found 72 with anomalies and 46 with normal vascular supply to the brain. In the insane he found 25 with anomalies and 7 with normal vessels. Is the difference between these greater than could be accounted for by chance? Granted, it looks fairly large. Yet a simple test shows that the probability of getting a difference as great as or greater than this by pure chance is 0.11. (The test used is Yate's modification of the Chi Square test.) Stopford has then an 11% chance of error in his conclusion, and this is more than the 5% standard. Since chance alone could account for the differences that he noted, he is not justified in assuming that sanity and insanity made any difference. This does not mean that the theory is not correct. It might be proven by further investigation on more subjects. On the other hand it might not, and if anyone is inclined to think that consideration of an 11% chance of error is hair-splitting accuracy, neither necessary nor desirable in such an investigation, he might well consider that Stopford's theory is not yet accepted, and indeed generally ignored.

Vascularity of Human Heart Valves:

D. Rutherford Dow and W. F. Harper in the Journal of Anatomy, July, 1932, discuss the vascularity of the valves in the human heart. They quote many conflicting reports concerning the presence or absence of vessels, especially in the mitral and tricuspid valves. They mention the theory that the number of vessels is greater in early life and decreases as valvular muscle disappears. They show too that this determination of valvular vascularity has a definite bearing on the theories regarding methods of bacterial spread in endocarditis. One school maintains that the bacteria gain foothold from the blood in the chambers, and the other considers the vessels in the valves to be the portal of entry . Dow and Harper injected the vessels in 37 normal hearts, 17 from foetuses and children under 10 years, 20 from persons over 10 years. The following table shows the results:

	Mitral Valve		Tricuspid Valve	
	Vessels		Vessels	
	Present	Absent	Present	Absent
Hearts from Persons under 10 years	7	10	7	10
Hearts from Persons over 10 years.............	7	13	3	17

No tests were applied to show the significance of these results.

In the summary the writers state:

"Vessels were more frequently present in the tricuspid and mitral valves of hearts under ten years of age than in those past this decade."

They do not definitely state that these results favor the theory of decreasing vascularity in the older valves, but they do leave the reader with that impression. Calculation shows that differences as great as or greater than those found between the two age groups for the mitral valve could be expected by chance alone in 68% of cases, if numerous samples of 37 hearts were examined. One would certainly not be justified in considering that age was a factor in causing the difference when chance could thus account for it in much more than 5% of cases. The difference in vascularity of the tricuspid valve between the two age groups is not significant either. The chance of error here would be about 15.7%.

Breast Function and Carcinoma:

W. Nicolson and M. D. Berman in the Annals of Surgery, May, 1936, discuss the five year results in a series of cases of carcinoma of the breast. They present a great deal of statistical information without any analysis to determine the extent to which their results are generally applicable. Their information regarding the effect of pregnancy and lactation is interesting, as are also the conclusions drawn from it.

Effect of Lactation in 205 Cases:

No. of	Cases Examined		Primarily Operable		5 Year Cures, all cases	
Lactations	No.	%	No.	%	No.	%
0	60	30	17	28.8	7	11.86
1	29	14	10	35.7	3	10.7
2	27	13.5	11	40.74	7	26
3	33	16.5	10	30.30	7	21.21
4	17	8.5	4	23.5	2	12
5	12	6	5	41.66	2	16.66
6	10	5	4	44.44	2	22.22
7	3	1.5	0	0	0	0
8	4	2	1	25	1	25
9	5	2.5	3	60	1	20
10	4	2	0	0	0	0
11	1	0.5	1	100	1	100

N.B.—The actual number of cases in each lactation (column 2) was calculated from the percentages and added to the original table.

In discussing the above table the writers state:

> "A larger percentage of cases occurred in the nulliparous breast than in any one group, but there were over three times as many in those which had lactated."

Sixty were nulliparous and one hundred and forty-five had lactated, over twice as many, but not over three times. The writers may have meant that less than 1/3 of all cases were nulliparous.

Quoting further:

> "The primary operability seems to increase with two or more pregnancies, as well as the percentage of five year cures. From this it might well be deduced that when a breast has performed the function for which it was intended there is less chance of a carcinoma being inoperable and a greater chance of cure."

These conclusions are based on a mere inspection of the percentages of primarily operable cases and five year cures in each class. But percentages can be very deceiving. Note that the larger percentages occur generally in classes that had fewer cases. How can one know that 100% based on one case or 44% based on ten cases differs materially from 28% based on sixty cases? Let us compare the actual number of primarily operable cases that were nulliparous or had one lactation with the number that had two or more lactations.

No. of Lactations	No. Primarily Operable	No. not Primarily Operable
One or none	27	62
Two or more	39	77

Analysis of these figures shows that the difference could easily be accounted for by chance alone, so it is not reasonable to suppose lactation was responsible. Differences as great as or greater than this could be expected by chance in 61% of cases. Comparison of the actual numbers of five year cures shows the following:

No. of Lactations	No. 5 Year Cures	No. Not Cured
One or none	10	79
Two or more	23	93

There is a 10% chance of finding differences as great as or greater than this in one series by chance alone, so it is not proven that number of lactations has any effect. The theory concerning the effect of breast function may be proven by further investigation, but is certainly not adequately proven by this series.

In conclusion—some knowledge of problems of random sampling and chance variation between samples is of great importance to the scientific investigator who wishes to select and classify his material so as to get most accurate results, and who wishes to make some general application of these results. Some such knowledge should be a part of the stock-in-trade of the medical man who must every day, consciously or unconsciously, take into consideration the numerous chance variations in the progress of any disease—variations due to a multiplicity of factors, mostly unknown. He must depend very often on his knowledge of probabilities too in the diagnosis, prognosis and treatment of a disease. Finally, a grasp of the same problems is necessary for the reader of most present day medical literature—unless he is willing to believe every opinion he reads. It gives real satisfaction to an individual to be able to make an intelligent estimate of the validity of a theory on the strength of the evidence submitted without being forced either to accept another's conclusions as fact or to cover complete ignorance by disregarding all the evidence with a sneering remark concerning statistics and lies.

BIBLIOGRAPHY

DOW, D. R. and HARPER, W. F., 1932, "The Vascularity of the Valves of the Human Heart", Journ. Anat., vol. 66, p. 610.

FISHER, R. A., "Statistical Methods for Research Workers", London; Oliver and Boyd.

MAINLAND, D., 1936, "Problems of Chance in Clinical Work", Brit. Med. Journ. Aug. 1, vol. 2, p. 221.

NICOLSON, W. P., and Berman, M. D., 1936, "Carcinoma of the Breast", Annals of Surg., vol. 103, No. 5, p. 683.

STOPFORD, J. S. B., 1936, "The Arteries of the Pons and Medulla Oblongata", J. Anat. Physiol, vol. 50, p. 131.

YULE, G. Udny, "An Introduction to the Theory of Statistics", 1929. London: Charles Griffin & Co.