trophy of the folds of the upper third of the stomach with an irregularity on the lesser curvature—the radiologist suggested a gastritis with perhaps a small ulcer. While the psychological aspects of the case were obvious on talking to the patient, the symptoms and inconclusive x-ray report raised questions of grave import. Gastroscopy showed a severe superficial gastritis with numerous gastritic ulcers, but the changes were chiefly in the distal parts of the stomach, and there was no hypertrophy of the folds, so that the x-ray report suggesting gastritis did not correspond either in position or character to what was seen in the mucous membrane.

Case 4.—A man aged 48 was seen in consultation with Dr. P. W. Tait of Sutton Coldfield on October 4, 1938. He had had a digestive upset in 1937, which began with complete loss of appetite, vomiting after food, and a sense of dull irritation in the epigastrium. The condition cleared up in two weeks, but a radiograph showed a normal gall-bladder, with a duodenal deformity suggesting an ulcer. He had a similar attack, immediately after the crisis, during which I saw him. I had difficulty in accepting his symptoms as being due to a duodenal ulcer, and with Case 3 fresh in my mind I made a diagnosis of gastritis. I was asked to see him again in March, 1939, as he had had three or four slighter attacks of the same type as before, during one of which he had melaena. He was investigated thoroughly in view of these events; a fractional test meal showed a very high free acidity, and a radiograph revealed a marked duodenal deformity, but gastroscopy failed to show any abnormality in the stomach; there were no signs of gastritis. At the operation by Mr. H. H. Sampson a very chronic and adherent duodenal ulcer was found and dealt with appropriately.

Cases 3 and 4 illustrate both the psychogenic basis of some dyspepsias and the fallacies in the way of making a clinical or radiological diagnosis of gastritis. Case 3 is especially interesting and recalls the work of Cullinan (1938) and Wittkower (1938) on the psychogenic factor in ulcerative colitis. It is unfortunately outside the scope of this paper to pursue the clinical aspects further, but the few illustrative cases may vivify the monotonous series of tables.

Summary and Conclusions

In 1,575 "gastro-duodenal" cases, culled from nearly 10,000 new out-patients, an indefinite diagnosis was made in 38 per cent.

Using the Wolf-Schindler gastroscope, 147 patients with radiologically negative or inconclusive dyspepsia have been studied; the series is almost consecutive.

The gastroscope has proved to be of great value in the routine investigation of this type of dyspepsia, and in about one-half of the patients an organic lesion of the stomach has been seen: usually a chronic gastritis of the superficial or hypertrophic types, atrophic gastritis being much less common.

Chronic gastritis can only be diagnosed with certainty by gastroscopy, and four case histories are given to emphasize the difficulties of clinical diagnosis.

REFERENCES

- Bulmer, E. (1938). *Practitioner*, **140**, 135.
Cullinan, E. R. (1938). *British Medical Journal*, **2**, 1351.
Schindler, R. (1937). *Gastroscopy*, p. 179, Chicago.
— and Ortmayer, M. (1936). *Arch. intern. med.*, **57**, 959.
Wittkower, E. (1938). *British Medical Journal*, **2**, 1356.

The Association of Port Health Authorities of the British Isles has published a pamphlet on the prevention of infestation of ships by the bed-bug. The illustrations in this are reproduced from the Ministry of Health *Memorandum on the Bed-Bug and How to Deal with It*. Requests for copies should be addressed to the honorary secretary of the Association, City Hall, Cardiff.

PROBLEMS OF CHANCE IN MEDICINE AND RESEARCH*

BY

L. J. A. LOEWENTHAL, M.D., M.R.C.P.

AND

W. A. WILSON, M.B., B.S., F.S.S.

(Uganda Medical Service)

In 1936 Mainland published an interesting paper in which attention was drawn to the frequency with which problems of chance in clinical medicine were ignored or falsely estimated. He formulated a method by which the odds could be calculated, even in small series, for or against a certain eventuality. Our reasons for doubting the validity of this method are given under the heading of "Statistical Appendix," where they are open to the criticism of statisticians. It is our intention in this paper to present a more exact method of estimating significance in a form easily accessible to those who are not conversant with statistical methods.

Significance

The term "significance" is widely and loosely used by many writers; to statisticians, however, it means something very definite—namely, "What are the odds against this result being due solely to the workings of chance?" As a first illustration, let us take two series of 100 patients in an epidemic of, say, influenza. It is assumed that these series are equivalent in every known respect and that they are observed during the same period of the epidemic and are subjected to precisely the same treatment. The mortality in one series is 15; what mortality may we expect in the other? If 15 die in the other series the agreement is closer than one can reasonably expect; with mortalities of 13, 14, 16, or 17 no surprise would be felt. But what if 35 die in the second series? There would be little hesitation in looking for an unknown factor—a "cause" for this difference in mortalities. Now somewhere between 17 and 35 in the present example there must be drawn a borderline separating differences which would arise so frequently in sampling that we may ascribe them to chance, and differences which would occur so seldom that some other cause must be sought.

The type of problem we are considering, then, is one where we want to know the answer to the following question: "What are the odds against 20 or more, 21 or more, 22 or more, etc., deaths in the second series?"

When these odds become great—for example, 199 to 1—we assume that chance has had very little to do with the matter, and some other factor must be responsible for the difference. An arbitrary line must therefore be drawn somewhere, and odds in excess of this arbitrary value counted as "significant." This is precisely what is done in statistical tables: the term " $P = 0.05$ " simply means that the probability (P) of such an eventuality happening through the workings of chance is 5 in 100, or, in betting parlance, "19 to 1 against." This is the borderline arbitrarily made between chance differences and significant differences.

In the table which we give two limits of significance are used—0.05 and 0.005. These are equivalent to odds of 19 to 1 and 199 to 1 respectively. The smaller figure in each cell, which represents the lower odds, is the

* Published with the permission of the Director of Medical Services, Uganda.

minimum number necessary to make the two series significantly different, and differences equalling or exceeding the higher figure may be regarded as highly significant. Examples will be given later to make this quite clear. The main use of the table, which has been calculated

Table of Differences required at Levels $P = 0.05$ and $P = 0.005$

No. in Each Series	Smallest Eventuality																								
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	18	20	25	35	50	75			
4	4																								
5	4	4																							
6	4	4	4																						
7	4	5	6																						
8	4	5	5	4																					
9	4	5	5	4	4																				
10	4	5	5	5	5																				
11	4	5	5	5	5	5																			
12	4	5	5	5	5	5	5																		
13	4	5	5	5	5	5	5	5																	
14	4	5	5	5	5	5	5	5	5																
15	4	5	5	5	5	5	5	5	5	5															
18	5	5	5	5	5	5	5	5	5	5	5														
20	5	5	5	5	5	5	5	5	5	5	5	5													
30	5	5	5	5	5	5	5	5	5	5	5	5	5												
50	5	5	5	5	5	5	5	5	5	5	5	5	5	5											
100	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5										
200	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5									
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	18	20	25	35	50	75			
Smallest Eventuality																									

Example of the use of the above table :

Series	I	II
Number	10	10
(that is, "10" in left-hand column)		
Deaths	2	5
Recoveries	8	5

Look up smallest eventuality (2) in top line. Corresponding cell opposite series 10 gives "5" as minimum difference necessary for significance. The actual difference between deaths was only 3.

by the exact factorial method (Fisher, 1934), is for small series. Where there are several hundred cases in each series the old methods, based on the standard error or Pearson and Elderton's χ^2 test, are satisfactory; but where series are unavoidably small they become unreliable. It may be impossible to ignore results from these small series and wait for larger numbers, for the rareness of a disease or the prohibitive cost of experimental animals may preclude this solution. The table is therefore presented in the hope that it will be of use to workers whose series are small and who are unable to make their calculations according to the somewhat laborious factorial method.

Use of the Table

As it stands the table assumes that there is an equal number of cases in the two series to be compared. The numbers in the left-hand column refer to the number of cases in each series. Along the top of the table is given the smallest eventuality in both series. Two figures are included in each cell in the body of the table: the smaller gives the difference which is required to make the odds 19 to 1—that is, $P = 0.05$; the larger gives the difference required for higher significance—that is, odds of 199 to 1, $P = 0.005$.

Example I.—Two groups of 10 guinea-pigs are inoculated with a suspension of streptococci; 2 in the treated group and 8 in the untreated group die. The difference is therefore 6. Look down the left-hand column for 10 (number in each series) and along the top for 2 (smallest eventuality). The figures in the appropriate cell are 5 and 7. The observed difference of 6 is therefore well within the 0.05 limit, and hence is significant, but cannot be regarded as *highly* significant.

Example II.—In two similar groups of 10 the deaths are 4 and 8. The difference is 4. It will be noticed that there is no entry opposite 10 to correspond with 4 (smallest eventuality) in the top line. This difficulty is easily overcome, for 8 and 4 deaths out of 10 respectively mean 2 and 6 recoveries. The smallest eventuality is therefore 2, and on referring to the appropriate cell it will be seen that a difference of 5 deaths would have been necessary for significance. In the example quoted the difference was only 4.

Example III.—The first illustration gave a mortality of 15 in a series of 100 patients. What mortality would be required to make the second series differ significantly? It is assumed that the mortality in the second series is higher, so that 15 is still the smallest eventuality. 100 in the left-hand column and 15 in the top row give figures of 11 and 17. To be highly significant there must be 17 more deaths in the second series—that is, 32. A difference of 11 or more, however, would still have been enough for most practical purposes.

Series containing Unequal Numbers

Obviously it is not always possible to have two series of precisely the same number for comparison, and tables to cover all possible combinations would be unwieldy. In many cases it is still possible to use the accompanying table in order to obtain an approximate idea of the significance of the difference observed in unequal series. From our own experience we should say that results obtained by the modifications given below are sufficiently near accuracy to enable, say, readers of medical journals rapidly to assess the value of numerical data presented to them.

1. Reduce the larger series to make its total equal that of the smaller, and reduce the deaths, etc., proportionately. On consulting the table, if the difference is now statistically significant, it must have been still more significant in the original data before reduction.

2. If the difference, after reduction of the larger series, is not quite significant the following may be tried: *providing that one series is not more than twice as large as the other* the arithmetic mean between the two series is found, and the figures respectively raised and reduced in proportion. The table may then be consulted.

Example :

Number in series	20	10
Deaths	4	7

The arithmetic mean between 20 and 10 is obviously 15; 4 reduced proportionately = 3. 7 raised proportionately

becomes 10.5. The difference is thus greater than 7, which from the table (15 in series; smallest eventuality 3) is ample for significance.

3. If one series is more than twice as large as the other the harmonic mean gives a better approximation.

Some Examples from Recent Literature

We propose here to cite four examples of actual figures published in medical journals, partly to provide further exercises for readers who will require to use the table in their own work and partly as an illustration of the inadvisability of drawing conclusions from data without subjecting such data to statistical analysis.

(A) Schutze (1939) gave the following figures (Table III in his paper):

Mice inoculated with Plague		
Treatment	Number	Survivors
M & B 693	6	2
Sulphone	6	4
Controls	6	0

On referring to the table (series = 6; smallest eventuality = 0) we find a difference of 4 necessary to establish statistical significance. In this case we may infer that sulphone has had an effect, but we may *not* conclude that M & B 693 had any effect.

(B) Anderson and Dowdeswell (1939):

Treatment of Pneumonia		
Treatment	No. of Cases	Deaths
M & B 693	50	1
Controls	50	8

The difference is thus 7 in series of 50, with a smallest eventuality of 1. The table shows that a difference of even 6 would have been significant at $P = 0.05$.

(C) H. Loewenthal (1939) (Table I in the paper under reference):

Infected Mice treated with Immune and Normal Sera in Addition to Sulphanilamide

Treatment	No. of Series	Deaths
Immune serum	4	0
Normal serum	4	3

The table shows that the minimum required difference for series of 4 is 4—that is, all must die in one group and all must recover in the other. Thus the experiment, though suggestive, cannot justify the author's statement: "There can be no doubt that . . . the combination of drug and (immune) serum is obviously superior."

(D) Cramer and Horning (1939) (Table I in authors' paper):

Adrenalectomy in Mice.—Sex Incidence of Mortality with and without Oestrinization

	Sex	No. in Series	Deaths
Adrenalectomy alone	M	8	4
	F	16	7
Adrenalectomy + oestrinization ..	M	19	4
	F	13	2
Combined	M	27	8
	F	29	9

The authors state: "There is a difference between the two sexes in so far as the females showed a better survival than the males in both groups." The table, of course, shows that there is no significant difference in these death rates. It can be demonstrated that the odds are 9 to 1 in favour of the differences in the combined series being exceeded by chance alone!

Summary

In small experimental and clinical series the assessment of the significance of observed differences is often a matter of difficulty.

We present a table by means of which the clinician and research worker can easily decide whether his results are significant or not.

The table, which is based on the factorial method, is so constructed as to permit quick reference and the accurate acceptance or rejection of data.

Statistical Appendix

The calculation of odds in problems similar to the ones we have dealt with above can be attempted in different ways. Historically the earliest approach was through the binomial distribution and the use of the well-known $\sqrt{npq}$ formula. This, however, fails when applied to small samples; an extension of its use has been suggested by Mainland (1936), where a confidence interval is calculated on the basis of a known probability p . In the type of experiment we are dealing with the probability has to be estimated from the data themselves, and Mainland's method would in our opinion require adjustment.

This brings us to the next method of approach—the test. The test, originally proposed by Pearson, the logical use of which was first defined by Fisher, has been successively amended (a) by Yates's correction for continuity and (b) by the substitution (Fisher and Yates, 1938) for χ^2 of varying values of χ ; their table, however, does not include some of the lowest frequencies which are often met with in clinical problems.

There are two factorial methods: Greenwood (1913) published tables of "the percentage frequency of successes in a second sample m after drawing p successes in a first sample n ." This at first sight seems identical with the type of problem before us, and the results obtained from his tables agree fairly closely with ours, as calculated by the exact factorial method given by Fisher (1934). We prefer the latter method.

In our table, columns 0 to 20 have been calculated by the exact factorial method; the others were deduced from the $\sqrt{npq}$ formula, and in some cases have also been checked by the factorial method. In addition, random entries have been checked by the Fisher and Yates χ^2 method, and agreement has invariably been obtained.

REFERENCES

- Anderson, T. F., and Dowdeswell, R. M. (1939). *Lancet*, 1, 252.
 Cramer, W., and Horning, E. S. (1939). *Ibid.*, 193.
 Fisher, R. A. (1934). *Statistical Methods for Research Workers*, 5th ed., Sect. 21.02, Edinburgh.
 — and Yates, F. (1938). *Statistical Tables for Biological, Agricultural, and Medical Research*, p. 37, Edinburgh.
 Greenwood, M. (1913). *Biometrika*, 9, 69.
 Loewenthal, H. (1939). *Lancet*, 1, 197.
 Mainland, D. (1936). *British Medical Journal*, 2, 221.
 Schutze, H. (1939). *Lancet*, 1, 266.

The College of Physicians of Philadelphia has awarded the Alvarenga prize to Dr. Harry Goldblatt, professor of experimental pathology and associate director of the Institute of Pathology, School of Medicine, Western Reserve University, Cleveland, Ohio, for his outstanding contributions to the pathogenesis of hypertension. This prize was established by the will of Pedro Francisco daCosta Alvarenga of Lisbon, Portugal, an associate fellow of the College of Physicians, to be awarded annually by the College of Physicians on each anniversary of the death of the testator, July 14, 1883, to the author of the best memorial upon any branch of medicine which may be deemed worthy of the prize.