Section of Epidemiology and State Medicine.

June 4, 1915.

Dr. W. H. Hamer, Vice-President of the Section, in the Chair.

The Statistics of Anti-typhoid and Anti-cholera Inoculations, and the Interpretation of such Statistics in general.

By Mr. Major Greenwood, jun., and Mr. G. Udny Yule.

Introduction.

Hardly any subjects within the range of preventive medicine are of more immediate importance than the methods of prophylaxis which ought to be adopted with respect to typhoid fever and cholera.

Typhoid fever has already been responsible for much illness and many deaths in nearly all the armies on active service, while cholera has taken toll of one at least of our enemies and one of our allies. Further, our troops are now fighting in a part of Europe and Asia which has always been a favourable soil for the development of epidemic cholera and was recently the scene of outbreaks among troops actually engaged in the present war.

Amongst the measures of prophylaxis which need to be discussed, that of preventive inoculation is clearly of exceptional interest, since such other hygienic precautions as experience would sanction are difficult, if not impossible, to carry out under service conditions. The expediency of inoculating troops against typhoid has been keenly debated, particularly since the outbreak of war, and the verdict of the medical profession has been almost unanimous. We have no evidence to bring forward which has not already done duty in the lay or scientific press within the last few months, and the remarks we shall have to make upon this branch of our subject will be brief. The data relating to cholera, on the other hand, have not attracted so much public attention, although, for the reasons just adduced, their importance is

great, and we shall be obliged to devote a good deal of space to their consideration. We have also been led to discuss various theoretical problems which might have been thought more suitable to the pages of a purely statistical journal. We are, however, satisfied that these questions of method ought to be studied in connexion with the practical problems from which they originate, and we hope that the arrangement and subject-matter will be found to conform to the logical necessities of the case.

In Section I we shall lay down the conditions which are, in our opinion, necessary to secure data from which valid statistical conclusions can be drawn and shall deal with the question of errors due to the fluctuations of sampling.

In Section II the available data are examined from the point of view defined in the first section.

In Section III we develop a statistical theory of the way in which immunization results can be interpreted.

In Section IV we consider the application of certain statistical methods of measuring association to data of the present kind.

Section I: The Conditions necessary for Valid Inference.

In the vast majority of cases one is provided with data arranged, or capable of being arranged, in the following manner: (*a*) A number of persons who have been inoculated against a certain disease and have not contracted the disease when exposed to risk; (*b*) a number of persons who have been inoculated and have contracted the disease; (*c*) a number of persons who have not been inoculated and have not contracted the disease; (*d*) a number of persons who have not been inoculated and have contracted the disease. If the statistics refer not to incidence but to fatality, the arrangement is the same, substituting "died of" or "recovered from" for "contracted the disease" or "did not contract the disease."

Given such data, the two following questions are those one is asked to answer: (1) Is there a *significant* difference between the attack (or fatality) rates of the two classes—i.e., is the observed difference greater than we could fairly attribute to the action of chance? (2) Assuming that the answer to (1) is in the affirmative, what is the degree of association between being inoculated and escaping from or being affected to a less degree by the disease, and how can we compare the result of one trial with that of another from this point of view?

It is not necessary at the moment to inquire whether these questions can be satisfactorily answered at all; whether they can be so or not is a question, principally but not entirely, of method, of statistical method, and its discussion presupposes the removal of certain material difficulties, assumes, in fact, that the four categories are homogeneous except as regards inoculation and death or attack. All the following conditions should be fulfilled by the data:—

(1) The persons must be, *in all material respects*, alike. The inoculated subjects must not differ in age, sex, social or racial constitution from the uninoculated subjects, unless it can be shown, on other grounds, that such differences do not affect the liability to contract the disease or, when one is investigating fatality, to die from it if contracted. We have italicized the proviso as to the similarity being material, for cases may be adduced in which although the persons are not *absolutely* alike they may be so in all material respects. For instance, in the case of typhoid, if personal attention to cleanliness under the ordinary conditions of army life can seriously reduce the liability to contract the disease, it might be argued that volunteers for inoculation are not homogeneous with uninoculated men, because the former, in virtue of their willingness to submit to inoculation, have proved themselves to be more careful than their comrades, and consequently would be expected to have less typhoid than the latter quite apart from inoculation. If, however, the ætiology of typhoid is such that no practicable care on the part of the individual can substantially reduce his risk of contracting the disease, then this particular heterogeneity is not material and can be disregarded.

In most cases, however, we have no such independent evidence, and any want of homogeneity in the above particulars must be deemed *prima facie* material.

(2) The effective exposure to the disease must be identical in the case of inoculated and uninoculated persons.

This condition, although clearly vital, has been often disregarded by medical writers. Evidently, if the uninoculated have been longer at risk than the inoculated, as would be the case if inoculation is carried on throughout an epidemic, and the statistics are compiled from the totals of inoculated and uninoculated at the end of the experience, the condition is not fulfilled. Neither is it fulfilled, even though the time of exposure is the same, if the conditions of exposure are not the same—e.g., if the records of inoculated persons are taken in one year and those of the uninoculated in another year. We could only

regard the heterogeneity of the second case as immaterial if a proof were furnished that the incidence (or fatality) of the disease is invariable from year to year; no such proof can be provided in the case of either typhoid or cholera.

(3) The criteria of the fact of inoculation and of the fact of the disease having occurred must be independent.

That the classification into inoculated and uninoculated must not be influenced in doubtful cases by knowledge of whether the disease had or had not been contracted may seem so obvious a stipulation that we need not trouble to set it out. In those cases in which the bacteriology of the disease is unequivocal and only a bacteriological diagnosis is admitted, while the fact of inoculation or its absence is recorded without fail in the case of every person exposed to risk, fulfilment of the condition only requires *bona fides* on the part of the observers. But when, as may quite frequently happen, doubts legitimately arise as to whether a given person had or had not been inoculated and the collection of records was entrusted to subordinates, convinced on other grounds of the efficacy of inoculation, it is easy to see how seriously biased statistics may be prepared without any evil intent. Such considerations no doubt afford any justification it may possess to Sticker's remark that "it is an open secret in India, as elsewhere, that for inoculation statistics the most favourable examples are always chosen" (Sticker, 1912, p. 313).

It will sometimes happen that the data themselves afford means of coming to a conclusion as to their material accuracy in the sense of the above remarks. This is so when we are provided with records of attack- or fatality-rates for inoculated and uninoculated persons in the case of several epidemics of the same disease. We shall return to this point in a later section, but may remark here that, upon almost any plausible theory of immunization, the attack- or fatality-rates of both classes in a series of epidemics of varying severity should be highly correlated. To take a simple illustration, the work of Dorset and others has shown that the larger the dose of virulent blood injected the larger must be the quantity of hyper-immune serum employed to protect a pig from swine fever, and that one of the difficulties of the method of simultaneous inoculation is to adjust the balance accurately. From this it follows that were the same dose of hyper-immune serum alone employed in different epidemics of the naturally occurring disease the fatality must vary among the inoculated in the same sense as among the uninoculated. This reasoning

is, we think, generally applicable, and if we find that over a series of epidemics the case- or fatality-rates of inoculated and uninoculated persons are not highly correlated, the discovery is evidence that the data are unreliable. Examples will be provided hereafter.

All the preceding conditions must be fulfilled to provide an accurate material basis for conclusions. But a further condition is necessary—namely, that the number of observations must be sufficient. We have, therefore, to deal with the first of our statistical problems —viz., to provide a criterion of the probability that any difference between the incidence- or fatality-rates of the inoculated and uninoculated persons actually found in the data should be deemed significant of an organic distinction between the two classes.

The criterion which has been adopted by most statisticians was introduced by Professor Karl Pearson in 1900 and was developed by him in the following way: If we know *a priori* that a table containing n "cells" should include m_1, m_2, m_3.....m_n, observations in the respective cells, then, if the standard errors of sampling and the correlation in errors of the various cells are of a certain form, the probability that fluctuations of sampling alone would give rise to a system of deviations $m_1 - m_1'$, $m_2 - m_2'$ ..... $m_n - m_n'$ where m_1', m_2', m_3' ..... m_n' are the actually observed contents of the "cells," or to any system of wider deviations, is a function of $\sum_1^n \frac{(m_n - m_n')^2}{m_n} = \chi^2$. The probabilities corresponding to different values of this function for various numbers of subdivisions or "cells" of the table have been computed by Elderton and are published in "Tables for Statisticians and Biometricians" (Cambridge University Press, 1914). Thus, if the number of subdivisions be 4 and χ^2 computed as above on the basis of some assumed theoretical distribution be 1, then the chance (P) is about 8 in 10 that errors of sampling would lead to as great a discrepancy as or a greater discrepancy than that actually observed between theory and observation. Had χ^2 been 2 the chance would have been about 57 in 100, had it been 8, the chance would have been about 46 in 1,000, and the probability goes on diminishing as χ^2 increases, being less than one in a thousand for $\chi^2 = 17$, and only one in a million for $\chi^2 = 30$.

The application of this test in the present case presupposes a theoretical distribution with which we compare our observed distribution, and, since we wish to determine the probability that any given arrangement has been brought about not by a fundamental connexion between the fact of inoculation and the fact of recovering or escaping

from the disease but by the operation of chance, we naturally assume as the theoretical law that the events are independent. In other words, if the chance of being inoculated is α and of escaping from the disease β, then in N trials we should expect to find N $\alpha\beta$ persons who have both been inoculated and escaped the disease; three similar expressions will give us the theoretical numbers falling within each of the other three subdivisions of the data, and then, since we know the actually observed frequencies within the four "cells," χ^2 can be computed and the required probability estimated with the aid of Elderton's table. It must be noticed that the application of this test to inoculation data is based upon an assumption. We do not, in fact, know the true values of α and β and must replace them by the observed ratios of the number of inoculated persons to the total frequency and the number of cases of disease (or deaths) to the total. This is not strictly correct, from the point of view of the general theory, but when we are dealing with such distributions as those actually in question, there is, perhaps, no more impropriety in making the assumption than in following the same course when we compute errors of simple sampling in the ordinary way.

A further point is worthy of remark. In our subsequent discussion we shall frequently compare the two ratios $\frac{\text{inoculated attacked}}{\text{all inoculated}}$ and $\frac{\text{uninoculated attacked}}{\text{all uninoculated}}$ which we may denote by p_1 and p_2. It might therefore be asked, why we should not adopt as our criterion of significance the ratio of $p_1 - p_2$ to its standard error, counting as significant all differences greater than some assigned multiple of the standard error. It will be found that if this plan is adopted deviations which, judged by the χ^2 test, are not improbable are much less likely to occur as the result of random sampling. This divergence between the results of the two tests is at first sight rather surprising and is not due to neglect of the correlation in errors between the subgroup frequencies. If the standard errors of p_1 and p_2 are worked out from first principles it will be found that the ordinary binomial form results, and that there is no correlation in errors between p_1 and p_2.

The explanation is, we think, as follows: The total number of distributions into, for example, four "cells" of n things which differ from the expected distribution by more than a certain margin is greater than the number of those which fulfil the further condition that the difference between the ratios $\frac{b}{a+b}$ and $\frac{d}{c+d}$ (p_1 and p_2 of our previous remarks, a being the number of inoculated who recovered, b the

number of inoculated who died, d and c corresponding frequencies of uninoculated) shall exceed a certain magnitude. The result will be that the probability of any arrangement having arisen from random sampling, or of any less probable arrangement, will be greater when estimated by the χ^2 method than when the other test is applied. It can, of course, be urged that we are really only concerned with the probability that chance might give rise to those arrangements which exhibit a difference between $\frac{b}{a+b}$ and $\frac{d}{c+d}$, and we are not convinced that the objection is invalid. We think, however, that the point is not free from difficulty and merits further consideration from the theoretical side, to which we have no space in the present paper to devote. Again, it is evident from our preceding remarks that any distribution deemed improbable on the basis of the χ^2 test will, *a fortiori*, be rejected by the other test. Consequently, in judging data by the χ^2 test we shall certainly not attach significance to results which might be the mere effects of random sampling and shall err on the side of caution.

Section II: Examination of the Available Data.

We shall now consider the statistics actually available for our inquiry, beginning with—

(*a*) *Anti-typhoid Inoculation.*

The only statistics of importance are still those compiled by the Anti-typhoid Inoculation Committee; the information afforded by the authorities with respect to inoculation of the troops in France and Belgium is far too meagre to be noticed here. The material criticisms to which the Committee's data were exposed are: (1) The inoculated men volunteered, they were not selected at random; (2) paratyphoid cases were excluded. With respect to the second point, we, having the fear of Dr. Hamer before our eyes, do not venture to express a personal opinion, but we may submit that the propriety of the exclusion has been admitted by all bacteriologists. We have already noticed (1) and need merely add that a study of the special conditions under which the test units were placed and of the ætiology of typhoid in the regiments confirms us in holding that the heterogeneity present was immaterial. A statistical question which arises is as to whether we should class as inoculated those who were so at the date of the last return made or only those actually inoculated at the time of arrival on the foreign

station. In the former case, we may be exaggerating the number of men who were inoculated during the whole period of exposure to infection, in the latter we shall underestimate it because many inoculations were done shortly after arrival.

The statistics have been arranged in both ways (Tables I and II) and the values of χ^2 computed. Both tables exhibit divergences from the distributions which might be expected if inoculation were without influence upon the chance of contracting typhoid, and these divergences are very unlikely to occur as errors of sampling. We may remark that if Professor Pearson's coefficient for a fourfold table be deemed an appropriate measure (see our discussion of this matter in Section IV), it will be found that the same conclusion emerges. This coefficient for Table I is 0·39 and for Table II 0·24, and each coefficient is several times the size of its standard error.

TABLE I.—ANTI-TYPHOID COMMITTEE'S DATA.

First arrangement.

	Not attacked	Attacked	Total
Inoculated	10,322 ...	56 ...	10,378
Not inoculated	8,664 ...	272 ...	8,936
Total	18,986 ...	328 ...	19,314

$\chi^2 = 180{\cdot}38$. P = less than 0·0001.

TABLE II.—ANTI-TYPHOID COMMITTEE'S DATA.

Second arrangement.

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

$\chi^2 = 56{\cdot}23$. P = less than 0·0001.

In view of the fact that the risk of serious injury resulting from inoculation is inappreciable and that even transitory discomfort is uncommon, provided the precautions as to rest and abstinence from alcohol immediately after the operation advised by the Committee are taken, the case in favour of anti-typhoid inoculation as a practical measure is very strong. We regret, however, that we have no more evidence to bring forward.

(*b*) *Anti-cholera Inoculation.*

A general account of the earlier attempts to confer immunity from cholera by inoculation will be found in Sticker's treatise (1912, pp. 307, &c.), and a fuller description in Haffkine's recent monograph. These accounts supplement one another, since Sticker views the subject from the standpoint of a convinced opponent of accepted teaching as to the pathology and epidemiology of cholera, while Haffkine writes with the natural enthusiasm of a pioneer. Haffkine's published experience relates mainly to the years 1894-99 and does not cover epidemics among troops on active service. More recent experience comparable with that of Haffkine is recorded by Nijland in the case of the Dutch East Indies and by Murata, who deals with some outbreaks in Japan.

The epidemiological history of cholera which is of most importance at the present time is that of the last Balkan War. This campaign called forth papers describing the conditions in the combatant armies which will repay examination. The Turks do not appear to have resorted to inoculation, indeed, so far as we can judge, their sanitary organisation was in all respects chaotic. Cholera broke out in their army after its defeat at and flight from Lule Burgas (October 30-31, 1912), (Simond, Pasteur Vallery-Radot, Kiamil Bey, Asseo). The ravages of the disease were enormous ; we have no trustworthy statistics, and accuracy of statement is not a *very* strong point of the Turks, but some observers put the number of deaths within a short period as high as 30,000. If this statement is even approximately correct, the gravity of the epidemic from every point of view need not be emphasized. This was less than three years ago, and our own troops are fighting in the same part of Europe and opposed by the very army which suffered this calamity. On the other hand, we have accounts of cholera in the Greek army which paint a very different picture (Savas, Cardamatis, Moutouses). It will be convenient to defer consideration of this outbreak until we come to the analysis of the statistical data derived from it.

We shall now examine the evidence provided by the various observers, beginning with that of Haffkine.

Haffkine's Data.

The eight sets of data extracted from Haffkine's treatise relate to the period 1894-99, and the practice of this epoch differed essentially from that now in vogue, since living cultures were employed. Two

vaccines were used in some of the inoculations. Of these, vaccine "II" was a living virus brought by successive passages through guinea-pigs to a uniform and stable degree of virulence. Vaccine "I" was an attenuated virus which had been found to protect guinea-pigs from the local lesion observed to follow the inoculation of vaccine "II." Haffkine, however, having discovered that no such lesion was produced in man, discontinued the use of vaccine "I" after 1896.

TABLE III.—CALCUTTA, 1894-96. PERSONS EXPOSED FROM FIFTH TO FOUR HUNDRED AND SIXTEENTH DAY AFTER INOCULATION.

	Not attacked	Attacked	Total
Inoculated	276	3	279
Not inoculated	473	66	539
Total	749	69	818

$\chi^2 = 29{\cdot}70$. P = less than 0·0001.

TABLE IV.—1ST BATTALION EAST LANCASHIRE REGIMENT, 1894. VACCINES I AND II. EXPOSURE FOURTEEN MONTHS AFTER INOCULATION.

	Not attacked	Attacked	Total
Inoculated	115	18	133
Not inoculated	520	120	640
Total	635	138	773

$\chi^2 = 2{\cdot}04$. P = 0·5652.

TABLE V.—BRITISH TROOPS AT CAWNPORE, 1894. EXPOSURE THIRTEEN MONTHS AFTER INOCULATION.

	Not attacked	Attacked	Total
Inoculated	75	0	75
Not inoculated	778	19	797
Total	853	19	872

$\chi^2 = 1{\cdot}83$. P = 0·6113.

TABLE VI.—2ND BATTALION EAST MANCHESTER REGIMENT AT DINAPORE AND CAMP BETA, 1894.

	Not attacked	Attacked	Total
Inoculated	193	0	193
Not inoculated	723	6	729
Total	916	6	922

$\chi^2 = 1{\cdot}60$. P = 0·6639.

TABLE VII.—GYA JAIL, 1894. INOCULATIONS DURING AN EPIDEMIC. VACCINES I AND II.

	Not attacked	Attacked	Total
Inoculated	200	8	208
Not inoculated	182	20	202
Total	382	28	410

$\chi^2 = 5{\cdot}90$. P = 0·1176.

TABLE VIII.—DURBHANGA JAIL, 1896. INOCULATIONS DURING AN EPIDEMIC. VACCINE II.

	Not attacked	Attacked	Total
Inoculated	105	5	110
Not inoculated	88	11	99
Total	193	16	209

$\chi^2 = 3{\cdot}18$. P = 0·3682.

TABLE IX.—MARGHERITA COOLIES. INOCULATIONS DURING AN EPIDEMIC, 1895. VACCINES I AND II.

	Not attacked	Attacked	Total
Inoculated	192	4	196
Not inoculated	113	34	147
Total	305	38	343

$\chi^2 = 37{\cdot}92$. P = less than 0·0001.

TABLE X.—CACHAR TEA ESTATES COOLIES, 1895-96. VACCINE II.

	Not attacked	Attacked	Total
Inoculated	5,751	27	5,778
Not inoculated	6,351	198	6,549
Total	12,102	225	12,327

$\chi^2 = 111{\cdot}92$. P = less than 0·0001.

Since this particular method is no longer employed, any lengthy comment is unnecessary. It will be seen that three only of the trials, Calcutta, 1894, Margherita, and Cachar, gave values of χ^2 greater than six for which P = 0·11; of these Cachar furnishes the largest value of χ^2. The Cachar results, which were obtained by Powell, will be reconsidered in the section devoted to the latter's publications; while recognizing that we have not sufficient evidence to permit of forming a complete judgment as to whether the inoculated and uninoculated be truly comparable in other respects, we shall not be going too far if we assert that these three results establish a presumption in favour of the value of Haffkine's inoculation.

Powell's Data.

Powell used a twenty-five to thirty-six hours' culture of the vibrio; sterilized water was added to one-third height of the agar, the growth was washed off and suspended in water by shaking; the dose used for an adult was c.mm. The whole of Powell's results to December, 1899, are quoted by Haffkine and summarized above under the heading Cachar Tea Estates. In order to see whether there was any marked heterogeneity, so far as the statistics could reveal it, between the

different estates we have analysed some of the separate returns. It will be seen that in most cases they are not based upon large enough numbers to yield satisfactory results, so that we should rest our conclusions as to Powell's evidence upon the Cachar total. A perusal of Powell's paper rather inclines one to believe that his material is fairly homogeneous, so that the value of the result in the Cachar material is increased. It may be noted that in this case Professor Pearson's normal coefficient is large ($0{\cdot}42 \pm 0{\cdot}02$).

TABLE XI.—KARKURI.

	Not attacked	Attacked	Total
Inoculated	409	3	412
Not inoculated	174	8	182
Total	583	11	594

$\chi^2 = 9{\cdot}34$. P = 0·0256.

TABLE XII.—KALAIN.

	Not attacked	Attacked	Total
Inoculated	1,625	5	1,630
Not inoculated	1,022	11	1,033
Total	2,647	16	2,663

$\chi^2 = 6{\cdot}08$. P = 0·1084.

TABLE XIII.—KALAINCHERRA.

	Not attacked	Attacked	Total
Inoculated	191	0	191
Not inoculated	608	8	616
Total	799	8	807

$\chi^2 = 2{\cdot}51$. P = 0·4802.

TABLE XIV.—DEGUBBER.

	Not attacked	Attacked	Total
Inoculated	431	5	436
Not inoculated	291	9	300
Total	722	14	736

$\chi^2 = 3{\cdot}27$. P = 0·3565.

TABLE XV.—DUNA.

	Not attacked	Attacked	Total
Inoculated	54	5	59
Not inoculated	46	15	61
Total	100	20	120

$\chi^2 = 5{\cdot}61$. P = 0·1351.

TABLE XVI.—RIVER.

	Not attacked	Attacked	Total
Inoculated	212 ...	1 ...	213
Not inoculated	42 ...	1 ...	43
Total	254	2	256

$\chi^2 = 1{\cdot}59$. P = 0·6662.

Murata's Data.

Murata employed a devitalized vaccine, heat was the agent of destruction, a temperature of 60° C. The dose was 1 c.c. containing 2 mg. of bacteria, and apparently only one inoculation was given. Murata seems to have some doubt as to the homogeneity of his statistics (p. 607), and he remarks that the dose should probably have been not 1 but 2 c.c. In view of the fact that the dosage, assuming that but one inoculation was given, was, judged by recent standards, insufficient, and that the temperature of sterilization exceeds that now employed, a point which is now known to be of great importance in the case of typhoid immunization, we should hardly anticipate very satisfactory results. Statistical analysis confirms this view. Only two cases in the series survive the χ^2 test, and of these one shows a *lower* proportion of attacks among the uninoculated. We think, in view of the facts recited, this series proves nothing one way or the other.

TABLE XVII.—KOOBE.

	Not attacked	Attacked	Total
Inoculated	14,939 ...	20 ...	14,959
Not inoculated	243,328 ...	753 ...	244,081
Total	258,267 ...	773 ...	259,040

$\chi^2 = 14{\cdot}48$. P = 0·0024.

TABLE XVIII.—HIMEJI.

	Not attacked	Attacked	Total
Inoculated	2,596 ...	0 ...	2,596
Not inoculated	28,680 ...	15 ...	28,695
Total	31,276 ...	15 ...	31,291

$\chi^2 = 1{\cdot}36$. P = 0·7189.

TABLE XIX.—KAWABE.

	Not attacked	Attacked	Total
Inoculated	8,135 ...	7 ...	8,142
Not inoculated	66,117 ...	88 ...	66,205
Total	74,252 ...	95 ...	74,347

$\chi^2 = 1{\cdot}25$. P = 0·7440.

TABLE XX.—MUKO.

	Not attacked	Attacked	Total
Inoculated	2,440	0	2,440
Not inoculated	80,713	62	80,775
Total	83,153	62	83,215

$\chi^2 = 1{\cdot}87$. P = 0·6022.

TABLE XXI.—AKASKI.

	Not attacked	Attacked	Total
Inoculated	9,297	3	9,300
Not inoculated	60,074	52	60,126
Total	69,371	55	69,426

$\chi^2 = 2{\cdot}99$. P = 0·3934.

TABLE XXII.—KAKO.

	Not attacked	Attacked	Total
Inoculated	2,729	1	2,730
Not inoculated	54,885	10	54,895
Total	57,614	11	57,625

$\chi^2 = 0{\cdot}46$. P = 0·9086.

TABLE XXIII.—INNAMI.

	Not attacked	Attacked	Total
Inoculated	657	0	657
Not inoculated	49,944	8	49,952
Total	50,601	8	50,609

$\chi^2 = 0{\cdot}11$. P = 0·9781.

TABLE XXIV.—SHIKAMA.

	Not attacked	Attacked	Total
Inoculated	3,098	2	3,100
Not inoculated	90,540	48	90,588
Total	93,638	50	93,688

$\chi^2 = 0{\cdot}08$. P = 0·9841.

TABLE XXV.—IBO.

	Not attacked	Attacked	Total
Inoculated	9,587	3	9,590
Not inoculated	86,032	1	86,033
Total	95,619	4	95,623

$\chi^2 = 18{\cdot}71$. P = 0·0003.

TABLE XXVI.—HIGAMI.

	Not attacked	Attacked	Total
Inoculated	3,173	0	3,173
Not inoculated	74,471	1	74,472
Total	77,644	1	77,645

$\chi^2 = 0{\cdot}04$. P = 0·9921.

TABLE XXVII.—TSUNA.

	Not attacked	Attacked	Total
Inoculated	19,567	11	19,578
Not inoculated	99,414	49	99,463
Total	118,981	60	119,041

$\chi^2 = 0{\cdot}16$. P = 0·9682.

Nijland's Data.

These data refer to both native and European inhabitants of the Dutch colonies. The author explicitly draws attention to the uncertain value of the statistics of the natives (1911, pp. 476-77); he does not give details of the immunizing process adopted, but it seems probable that one injection of a living vaccine was employed. As will be seen from the tables the population figures are often round numbers and, as is clear from the author's account, must be only approximations. In most of the tables we have placed cases occurring within a few days of inoculation to the credit of the inoculated group, but in the tables of Europeans, 1911-12, we have accepted the author's classification, which, for instance, transfers a case (1912) in an inoculated person, beginning only two days after inoculation, to the uninoculated class. It will be seen that the majority of the tables yield values of χ^2 inconsistent with the hypothesis that the differences between the groups of inoculated and uninoculated persons are attributable to chance. The general sense of Nijland's data is clearly favourable to inoculation, but his paper gives us little confidence as to the worth of the native statistics, while those compiled for Europeans are not accompanied by details as to age, sex, and social class.

TABLE XXVIII.—SINGARADJA.

	Not attacked	Attacked	Total
Inoculated	4,087	5	4,092
Not inoculated	113,856	1,144	115,000
Total	117,943	1,149	119,092

$\chi^2 = 31{\cdot}49$. P = less than 0·0001.

TABLE XXIX.—SEMARANG.

	Not attacked	Attacked	Total
Inoculated ...	8,332	8	8,340
Not inoculated	84,444	556	85,000
Total	92,776	564	93,340

$\chi^2 = 39{\cdot}40$. P = less than 0·0001.

TABLE XXX.—DENPASSER.

	Not attacked	Attacked	Total
Inoculated ...	1,730	3	1,733
Not inoculated	153,549	451	154,000
Total	155,279	454	155,733

$\chi^2 = 0{\cdot}85$. P = 0·8311.

TABLE XXXI.—SELONG.

	Not attacked	Attacked	Total
Inoculated ...	4,870	5	4,875
Not inoculated	153,096	904	154,000
Total	157,966	909	158,875

$\chi^2 = 19{\cdot}50$. P = 0·0002.

TABLE XXXII.—EUROPEANS OF SEMARANG.

	Not attacked	Attacked	Total
Inoculated ...	2,450	0	2,450
Not inoculated	2,638	38	2,676
Total	5,088	38	5,126

$\chi^2 = 35{\cdot}05$. P = less than 0·0001.

TABLE XXXIII.—EUROPEANS IN SOERAKARTA.

	Not attacked	Attacked	Total
Inoculated ...	620	2	622
Not inoculated	814	13	827
Total	1,434	15	1,449

$\chi^2 = 5{\cdot}42$. P = 0·1465.

TABLE XXXIV.—EUROPEANS IN JAVA, 1910-11.

	Not attacked	Attacked	Total
Inoculated ...	3,999	1	4,000
Not inoculated	5,942	58	6,000
Total	9,941	59	10,000

$\chi^2 = 36{\cdot}28$. P = less than 0·0001.

TABLE XXXV.—EUROPEANS IN JAVA, 1912.

	Not attacked	Attacked	Total
Inoculated ...	6,999	1	7,000
Not inoculated	3,687	13	3,700
Total	10,686	14	10,700

$\chi^2 = 21{\cdot}05$. P = 0·0001.

Zabolotny's Data.

This author furnishes us with an interesting account of epidemic cholera in Russia during 1907-08, in which years the disease was widely prevalent. In the city of Petrograd the disease became epidemic in the month of August, 1908; during September of that year the maximum number of cases reached 419 per day, the total for the month being 6,799. By the middle of April, 1909, there were 10,311 cases with 4,006 deaths. The number of persons inoculated in Petrograd during the 1908 epidemic is said to have been more than 15,000, among whom twelve cases only occurred. The total population of the city was about a million and a half. During 1907 more than 4,000 persons were inoculated in the province of Astrachan, more than half of these twice, and the attack-rate was 5 per mille. A comparison is instituted between the inoculated and uninoculated employees of Nobel frères in the city of Tzaritzine (in 1908). The figures are: Of 590 inoculated one was attacked, of 2,390 not inoculated eighteen. These, and other results cited, certainly suggest that the inoculations were of benefit, but we are of opinion that Zabolotny's figures do not lend themselves to exact analysis. As we shall have occasion to demonstrate, à propos of the statistics relating to the Greek Army, inoculation returns compiled during the course of an epidemic are very difficult to interpret unless we have information as to the period of exposure to risk of those persons not inoculated before the disease appeared, but subsequently so protected. The information provided in Zabolotny's paper is insufficient to enable us to avoid the pitfalls, and we cannot, therefore, use his data.

Savas's Data.

The data now to be discussed are of such importance and are so certain to be quoted by all writers on the subject, that it will be necessary to consider them very carefully. Savas's experience relates to the Greek Army during the second Balkan War and the essential facts of the situation are as follows: During the first Balkan War, that between the Allied States and Turkey, no cases of cholera occurred in the Greek Army. As we have seen, however, there was a murderous outbreak among the Turkish troops, and the Bulgarians became infected before and during the siege of the Chataldja lines. From here they carried the infection into Thrace and Eastern Macedonia, territories containing many Greek inhabitants. Fugitive Greeks are assumed to

have conveyed the seeds of infection to within the Greek lines, the first case occurring May 11-24, 1913. At the outbreak of war the exigencies of the service prevented the complete inoculation of the Greek Army and the process was not, in fact, completed until the cessation of hostilities in July. While on the march, the troops drank water polluted with Bulgarian corpses, and civilian fugitives spread the disease somewhat widely through the towns and villages of Macedonia. Not only was there pollution of the water, but flies were abundant and it was noticed that when the troops were encamped at an altitude of more than 1,000 metres, where the streams were unpolluted and flies rare, the cases were rarer. As to the actual course of the epidemic, we are informed that the first case occurred in May, that in July the epidemic declined rapidly and that it came to an end early in October; weekly or monthly returns are not provided. Including cases among the inhabitants of Macedonia, it is estimated that there were from 5,200 to 5,300 attacks with 1,665 to 1,700 deaths. Among the troops there were 2,503 cases, with 515[1] deaths (it will be remembered that in the Turkish Army the deaths during a shorter period were estimated at some 30,000, but the estimate is uncertain and the numbers exposed to risk probably much larger).

From Savas's account, as well as that of his assistant Moutouses, it is plain that the sanitary organization of the Army was of a high order; efforts were made, apparently with success, to separate cholera patients from the wounded practically in the fighting line, and field laboratories were available for deciding the diagnosis. Much importance is assigned to the fact that the control of both civil and military patients was in the same hands.

The history of the inoculations is as follows: They were commenced about two months before the outbreak of war, and all the sanitary division had been inoculated before cholera appeared. The first and seventh army divisions were inoculated after the first case (that in May) had been diagnosed, and the vaccination of the whole army was complete at the end of July or the beginning of August; in all, about 150,000 men were inoculated. In addition a very large number of civilians in Salonika, Macedonia and Old Greece, some 500,000, were inoculated. The material for inoculation was derived from a culture of the vibrio isolated from the fæces of a Bulgarian, its virulence being exalted by

[1] These are the cases returned from the cholera ambulances and military hospital, but it seems probable that some were civilians, for in another table only 2,192 are allotted to the army divisions.

passage through guinea-pigs and maintained constant during the whole period of preparation. A bacillary emulsion containing 4 mg. of bacteria per cubic centimetre was employed for inoculation, it was sterilized at 56° C. and 5 per 1,000 carbolic was subsequently added. The dosage was ½ c.c. for adults, followed eight days later by 1 c.c. The inoculations were rarely attended by complications, but in a few cases fulminating cholera developed within a few hours of the first inoculation. Savas is disposed to regard this as evidence of the reality of a negative phase.

TABLE XXXVI.—MORBIDITY AND VACCINATION IN THE GREEK ARMY.

	PARTICULARS AS TO VACCINATION				CHOLERA CASES				
Division	Original strength	Vaccinated once	Vaccinated twice	Not vaccinated	Vaccinated once	Vaccinated twice	Not vaccinated	Doubtful	Total
I	16,259	993	11,526	3,740	48	56	31	—	135
II	12,439	66	22,199 [12,199]	174	66	23	174	96	359
III	11,922	3,138	8,472	312	147	260	48	—	455
IV	14,822	3,306	8,397	3,119	36	20	96	—	152
V	10,859	3,010	7,819	30	90	83	19	—	192
VI	11,081	3,690	6,182	1,209	36	39	119	—	194
VII	13,500	250 [258]	13,192	50	62	64	13	—	139
VIII	9,082	24	9,032	26	7	8	26	—	41
X	11,404	120	10,976	308	120	64	308	—	492
XI	3,435	6	3,426 [3,429]	—	6	27	—	—	33
Totals	114,803	(14,603)? 14,613* [14,611]	(101,221)? 91,224 [91,224]	8,968	618	644	834	96	2,192
Total morbidity	—	—	—	—	42 per 1,000	7 per 1,000	93 per 1,000	—	19 per 1,000

If the figures in the columns are correct the totals should be as marked "?" The figures in brackets are suggested as being more probably correct, and in that case the totals in square brackets will be the right ones. * This total is quoted in the text as 14,411.

In Tables XXXVI and XXXVII we give those of Savas's statistics which we propose to discuss. His statistics of fatalities are not, we think, suitable for analysis on the lines of this paper, being specially open

to the criticisms which we shall make with respect to the incidence records. We have noted on Table XXXVI certain minor discrepancies in the figures and indicated the corrections which seem necessary (*vide infra*).

TABLE XXXVII.—MORBIDITY OF THE SANITARY COMPANIES OF THE DIVISIONS.

Division	Number	Vaccinated once	Vaccinated twice	Not vaccinated	CHOLERA CASES			
					Vaccinated once	Vaccinated twice	Not vaccinated	Total
I	417	11	402	4	—	—	—	—
II	323	—	323	—	—	—	—	—
III	280	—	280	—	—	—	—	—
IV	284	—	284	—	—	—	—	—
V	241	—	241	—	—	4	—	4
VI	393	51	342	—	3	—	—	3
VII	280	—	280	—	—	2	—	2
VIII	308	—	308	—	—	—	—	—
X	371	—	371	—	—	4	—	4
Totals	2,897	62	2,831	4	3	10	—	13
Total morbidity	—	—	—	—	48·38 per 1,000	3·5 per 1,000	—	4 5 per 1,000

Before discussing the statistics themselves, there are some epidemiological points to notice. It is not clear whether we ought to regard the record as of a single outbreak beginning in May and remitting in July or as of a series of separate outbreaks. On the former hypothesis we cannot, in the absence of daily or weekly returns, say whether there was an explosive rise, followed by a decline at first rapid and then slower, when the numbers of fresh cases had fallen to a comparatively low level. Such a course plotted as a curve will give us two unequal areas (the abscissæ being units of time measured from the date of the first case and the ordinates—numbers of cases) if the division is made by a vertical line drawn through the point corresponding to the midpoint in time of the epidemic, and the number of cases during the first period will exceed that of the second period. This is the usual, but not

invariable, state of affairs in epidemic cholera (Table XXXVIII) observed in civil populations, and most of such epidemics have been attributed, on more or less satisfactory grounds, to water-borne infection. Polluted water was certainly one of the sources of infection in the Greek experience, but we have no reason to assume that this was the only or even the most important factor.[1] If, however, we suppose that the

TABLE XXXVIII.—DISTRIBUTION OF CASES OF CHOLERA IN VARIOUS EPIDEMICS AND SERIES OF EPIDEMICS (DATA FROM STICKER, 1912, p. 151).

The numbers of cases in each week of the epidemic were provided, and if there were an odd number of weeks, the cases of the middle week were divided equally between the first and second halves of the epidemic.

Epidemic	Prussia, 1848-59 (monthly)	Hamburg, 1831-73 (monthly)	Petrograd, 1831	Moscow, 1831	Berlin, 1831	Hamburg, 1831	Vienna, 1831	Prague, 1831	Lübeck, 1832	Rostock, 1832
Number of cases in the first period	193,146	12570·5	8370·5	5,204	1415·5	861	2,107	1,555	230 5	453·5
Number of cases in the second period	118,890	6823·5	434·5	1,047	772·5	77	1,380	1,831	906·5	226·5
Total ...	312,036	19394·0	8805·0	6,251	2188·0	938	3,487	3,386	1137·0	680·0
100 × number of cases in the first period divided by the total ...	61·9	64·8	95·1	83·3	64·7	91·8	60·4	45·9	20·3	66·7

epidemic was of the type described, some reflections suggest themselves. The explanation of the forms of epidemic curves is not definitely known (*see* Brownlee, Greenwood, Ross, Sticker), but the cause is only rarely the exhaustion of susceptible persons. It appears that from a certain point in time, often perhaps from the very commencement of the epidemic itself, the infectivity of the disease diminishes. If, therefore, we were to inject persons exposed to risk with coloured water, and

[1] It is to be remembered that, according to Eckert ("Die Rolle der Kontaktinfektion in der Epidemiologie der Cholera," *Berl. klin. Woch.*, 1913, 1, p. 2326), water-borne infection was not responsible for the outbreak of cholera among the Bulgarian troops before the Chataldja lines in the first Balkan War, an outbreak which cost them from 16,000 to 29,600 men and may quite possibly have saved Constantinople. Personal infection was thought to have played the chief part.

were so to plan our operations that the absolute number of persons inoculated in each unit of time, say each day or each week, was constant, and were finally to count up the numbers of inoculated persons and of attacks amongst them and prepare similar statistics of uninoculated persons, we should inevitably find that the attack-rate was lower in the former case. This would happen were the infectivity constant, and a declining infectivity would accentuate it.

We can illustrate the point readily. Let us suppose that an initial population of 1,000 is exposed to risk during a period of three weeks, that during each of these weeks the chance of acquiring a disease is one in a hundred and that at the beginning of each week fifty persons receive an injection of coloured water. At the end of the first week there will be 950 uninoculated and 50 inoculated persons, and ten cases of disease will have occurred—9·5 among the uninoculated, 0·5 among the inoculated. Stopping at this point, we have identical attack-rates in the two classes—viz., 1 per cent. At the beginning of the second week, we are left with 940·5 uninoculated and 49·5 inoculated men who have not had the disease; fifty of the former are now inoculated, so that the numbers become 890·5 uninoculated, 99·5 inoculated; subjecting all to the attack-rate of 1 per cent., 8·905 uninoculated and 0·995 inoculated fall victims. Adding up from the beginning we reach 100 inoculated with 1·495 cases and 900 uninoculated with 18·405 cases—rates of 15 and 20 per mille respectively. At the end of the third week we shall have 150 inoculated with 2·98 cases, and 850 uninoculated with 26·72 cases—rate per mille of 20 and 31; thus the relative advantage of the inoculated increases. This error, which is inherent in the method of summation adopted, would have been increased if the chance of infection instead of being constant had diminished as time passed —a phenomenon which we have seen reason to anticipate in practice. The conclusion is that, if we are only provided with the total number of inoculations performed during the epidemic and the allotment of attacks between inoculated and uninoculated classes, we are almost sure to find that the inoculated have an advantage and are by no means warranted in concluding that this is any more than a necessary consequence of the manner of compilation. Of course, the explanation of the fallacy is that the period of exposure to risk is not the same in the two classes; the men who would perhaps have been inoculated next week have a chance of acquiring the disease this week, and if they do so will naturally be counted as unvaccinated attacked persons. There is no way of circumventing

this error on the basis of summarizing tables and it must always, in greater or less degree, affect the statistics of inoculations performed during an epidemic. No doubt if full details as to time and place of inoculation and attack were furnished, we could isolate the true from the spurious advantage; but that, in the stress of active campaigning, such particulars are likely to be recorded is improbable. The conclusion seems to be that Savas's divisional statistics cannot be made the basis of sound reasoning with respect to the value of inoculation; it is possible, indeed, as we shall show directly, probable that inoculation did confer a benefit upon the troops, but we cannot express this numerically in a manner to be comparable with the typhoid results. A detailed inspection of Savas's figures only confirms the above conclusions. To begin with there are some arithmetical discrepancies. The table is copied accurately from Savas's paper, but the totals for once vaccinated and twice vaccinated persons are not correct—the sums of the columns are respectively 14,603 and 101,221, instead of 14,613 and 91,224. As in all but two instances the sums of the numbers once, twice, and not vaccinated agree with the numbers entered as the original strengths of the divisions, it is probable that for 22,199 twice vaccinated in the second division we should read 12,199: with this correction, the discrepancy is reduced to three in the case of the twice vaccinated and ten in the case of the once vaccinated, an unimportant difference. But there is some uncertainty about the true figures, for in the text the number of once vaccinated persons is given as 14,411.

Further inspections of the details reveal other peculiarities. In the case of the second division, every person once vaccinated and every person not vaccinated was attacked by cholera, and the same remark applies to the tenth division. In the eighth division every uninoculated man had cholera. Savas speaks so enthusiastically of the value of inoculation and lays so much stress on the experience of certain regiments that, had these divisional results really afforded a comparison as fair as it is striking, we do not doubt that he would have commented on them. The only conclusion we can draw is that all or the majority of the twice inoculated men in the two divisions cited became so after the uninoculated men were attacked, and this is perhaps what Savas means by saying that the second division was completely inoculated. For these reasons, we do not think the treatment adopted in the case of the other observers' material could deduce results of scientific value from Savas's divisional statistics.

This is disappointing because, among other things, if we could take the figures at their face value we could test certain statistical processes, since we have six instead of only four "cells" owing to the fact that there are records of twice, once, and not vaccinated men. We have, indeed, made a few calculations, but although they might have some interest from the theoretical side, the above-mentioned facts deprive them of practical value and we shall not reproduce the constants.

It will be seen, then, that the data of Table XXXVI do not of themselves add much to our knowledge of the value of inoculation, but combined with another set of observations (Table XXXVII) interesting results can be obtained. It will be remembered that one branch of the Greek Army—viz., the sanitary corps—was completely inoculated prior to the outbreak of cholera. The members of this corps were of the same social class as the combatants; they had indeed received a short course of sanitary instruction from the medical officers, but against this must be set the fact that they may have been more exposed to risk than ordinary soldiers; for instance, it was often their duty to carry cholera patients, sometimes on their backs and down mountain paths under conditions rendering it difficult or impossible to avoid contamination with fæces or vomit.

It seems reasonable to think that the sanitary service did not occupy a more but a less favourable position than the combatants as regards the risk of infection; the one particular in which they were distinguished from all or nearly all the combatants was, that practically every man had been twice vaccinated before the disease broke out. If now we construct a fourfold table (Table XXXIX), the divisions being

TABLE XXXIX.—SAVAS'S DATA.

	Not attacked	Attacked	Total
Sanitary corps	2,884	13	2,897
Combatants	112,613	2,192	114,805
Total	115,497	2,205	117,702

$\chi^2 = 32{\cdot}79$. P = less than 0·0001.

TABLE XL.—SAVAS'S DATA.

	Not attacked	Attacked	Total
Sanitary corps	2,884	13	2,897
Combatants	109,729	2,179	111,908
Total	112,613	2,192	114,805

$\chi^2 = 33{\cdot}85$. P = less than 0·0001.

between attacked and not attacked, as usual, and between combatants and sanitary corps, we find $\chi^2 = 32{\cdot}79$, so that the odds against the differences between the attack-rates being a mere chance event are very heavy. We must infer that the sanitary corps somehow or other acquired a higher degree of immunity than the combatants.[1] *If* we adopt Savas's view as to class of men and degree of exposure to risk, the only inference left is that the immunity was consequent upon early and complete inoculation. We have been at some pains to detect any source of material fallacy in this reasoning and have failed to do so. Assuming that it is correct, anti-cholera inoculation carried out with the technique likely to be used in the case of our own troops is a prophylactic step of importance, although an exact statistical measure of the degree of relative immunity conferred cannot be provided.

That anti-vaccinists will accept our conclusions respecting Savas's data is wildly improbable. We know no more about the Greek Sanitary Corps than Savas tells us. They may all have been vegetarians, or non-smokers, or red-headed, and all or any of these things may render them less likely to contract cholera, but we do not see why objections which no sensible man would allow to influence him in the affairs of ordinary life should suddenly acquire scientific importance when the question is one of interpreting statistics. Our conclusions, then, respecting the Greek experience are that, although no inference can be drawn from a comparison of the attack-rates upon inoculated and uninoculated soldiers in the combatant units, yet the striking difference between the incidence upon the sanitary corps and that upon the rest of the army is evidence in favour of the efficacy of the process.

Briefly summarizing our study of anti-cholera inoculation statistics, it appears that in several cases the difference between the inoculated and uninoculated in respect of cholera incidence is greater than can be attributed to the operation of chance. Were it permissible to combine the results, then the combined improbability of random sampling accounting for the deviations from the systems deduced from a hypothesis of strict independence would be great indeed.[2] Naturally, these conclusions presuppose the fulfilment of the material conditions laid

[1] It not being clear whether Table XXXVI does or does not include the data of Table XXXVII, we have repeated the calculation shown in Table XXXIX, assuming that the sanitary corps were already included in Table XXXVI. The difference is not of importance (Table XL).

[2] This combination has not been much used in practice where a series of tables for which the values of χ^2 are known is in question ; but see Pearson and Heron, p. 314.

down at the beginning, and in each set we have, to the best of our ability, indicated whether such was the case or not. The test of

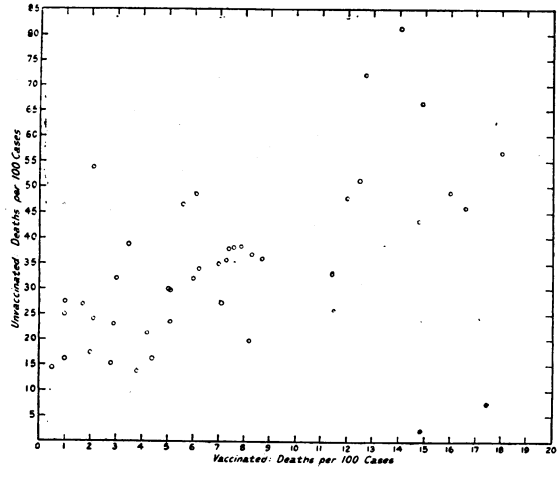

FIG. 1.

Correlation between small-pox case mortality for inoculated and uninoculated in various epidemics.

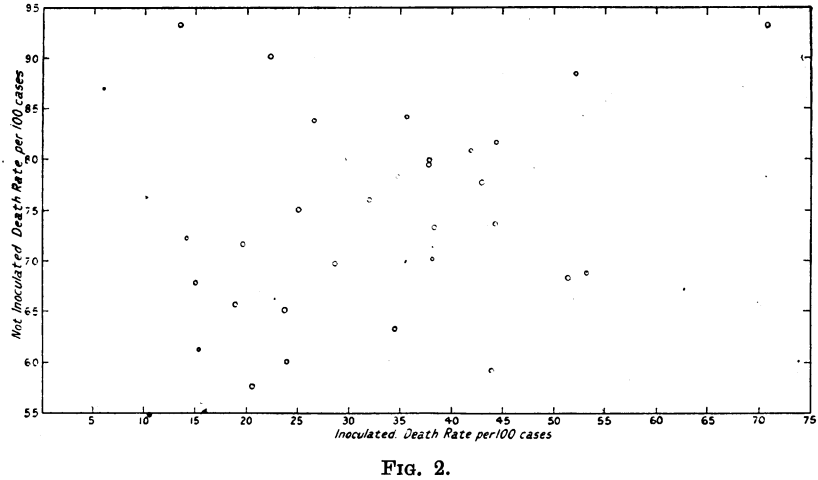

FIG. 2.

Correlation between plague case mortality for inoculated and uninoculated in various districts.

consistence to which we alluded (p. 116) cannot be satisfactorily applied to such short series and we ought to warn the reader that in the case

of the only two diseases for which we could obtain long series—i.e., many cases of groups of inoculated and uninoculated persons exposed to epidemics of different severity—the results are bad. This is illustrated in figs. 1 and 2. The first (Table XLI) refers to small-pox and

TABLE XLI.—SMALL-POX CASE MORTALITY. PERCENTAGE OF DEATHS AMONG THOSE ATTACKED ACCORDING TO WHETHER VACCINATED OR UNVACCINATED.

		Death-rate per 100 cases	
		Vaccinated	Unvaccinated
France	1816-41[1]	1·0	16·2
Marseilles	1828[1]	1·0	25·0
Würtemberg	1821-35[1]	7·1	27·3
,,	1840-50[1]	3·5	38·9
Canton Vaud	1825-9[1]	2·1	24·0
Bohemia	1835-55[1]	5·1	29·8
Milan	1830-51[1]	7·6	38·3
Verona	1828-39[1]	5·6	46.6
Breslau	1831-3[1]	2·1	53·8
Copenhagen	1828-37[1]	1·0	27·6
Vienna Hospital	1834[2]	12·5	51·2
,, ,,	1837-56[1]	5·0	30·0
,, ,,	1859[1]	3·8	13·8
,, ,,	1870[1]	2·0	17·4
Prague Children's Hospital	1840-58[1]	3·0	32·0
London Small-pox Hospital	1836-56[1]	7·0	35·0
,, ,, ,,	1863[1]	12·0	48·0
,, ,, ,,	1864[1]	8·7	36·0
,, ,, ,,	1865[1]	7·4	38·0
,, ,, ,,	1866[1]	7·3	35·7
,, ,, ,,	1867[1]	8·3	36·8
,, ,, ,,	1868[1]	6·2	34·0
,, ,, ,,	1870[1]	7·9	38·5
,, ,, ,,	1871[1]	14·9	66·5
Poland	1871[1]	11·4	33·0
Frankfort Town	1871[1]	16·0	49·0
,, ,,	1872[1]	16·6	46·0
,, Government 1 District	1871[1]	14·8	43·4
Liverpool Small-pox Hospital	1870[1]	12·7	72.0
Berlin Lazaretto	1871-2[1]	14·1	81·3
Coblentz	1871[1]	18·0	56·9
Quebec	1819-20[2]	1·7	27·0
Philadelphia	1825[2]	—	60·0
Darkehmen	1828-9[2]	—	18·8
Carniola	1834-5[2]	4·4	16·2
Carinthia	1834-5[2]	0·5	14·5
Adriatic	1835[2]	2·8	15·2
Lower Austria	1835[2]	11·5	25·8
Galicia	1836[2]	5·1	23·5
Dalmatia	1836[2]	8·2	19·6
Kiel	1852-3[2]	6·0	32·0
Malta	—— [2]	4·2	21·1
Epidemiological Society Returns	—— [2]	2·9	23·0
Illinois	—— [2]	6·1	48·6

[1] From Second Report of the Vaccination Committee, Appendix II, Table N (Dr. A. F. Hopkirk), from *Zeitschft. des K. Preuss. Stat. Bureau*, Jg. xiii, p. 155.

[2] From Sixth Report of the Vaccination Committee, Appendix X (Sir B. A. Whitelegge), p. 660. From article "Vaccination," by Dr. S. A. Abbott, in Wood's "Reference Handbook of the Medical Sciences." For references to originals *see* loc. cit.

the figures are taken from Appendix X of the Sixth Report of the Royal Commission on Vaccination and from Appendix II of the Second Report. The second (Table XLII) is compiled from returns of plague inocula-

TABLE XLII.—PLAGUE: CASE MORTALITY, AMONG INOCULATED AND UNINOCULATED PERSONS.[1]

		INOCULATED			UNINOCULATED		
		Attacks	Deaths	Death-rate per 100 cases	Attacks	Deaths	Death-rate per 100 cases
Kirkee	1897	32	17	53·1	143	98	68·5
Dharwar ...	1898	129	54	41·9	1,100	889	80·8
Gadag	,,	193	83	43·0	278	216	77·7
Belgaum ...	,,	78	40	51·3	506	346	68·4
Ahmednagar ...	,,	70	31	44·3	563	415	73·7
Aden	1900	23	8	34·8	83	65	78·3
Nagpur	1906	25	8	32·0	204	155	76·0
Kollegal ...	1908	68	18	26·5	136	114	83·8
Kunagalli ...	,,	45	20	44·4	49	40	81·6
Mudigandam ...	,,	59	21	35·6	38	32	84·2
Madanhalli and Haranapuram	,,	135	51	37·8	234	186	79·5
Palayam ...	,,	73	28	38·4	60	44	73·3
Dharampur ...	1910	24	6	25·0	209	157	75·1
Gujranwala ...	,,	41	18	43·9	1,013	600	59·2
Amritsar District	,,	90	17	18·9	2,513	1,651	65·7
Neemuch ...	,,	39	8	20 5	382	220	57·6
Nagpur Mills ...	,,	74	10	13·5	7,770	7,253	93·3
Yeotmal ...	,,	27	6	22·2	407	367	90·2
Harihar... ...	,,	92	18	19·6	283	203	71·7
Palni	,,	89	63	70·8	179	167	93·3
Bijapur	1911	28	8	28·6	1,228	857	69·8
Gadag	,,	20	3	15·0	1,695	1,150	67·8
Salem	,,	434	226	52·1	1,693	1,495	88·3
Channapatna ...	,,	71	10	14·1	166	120	72·3
Kirkee	1912	21	8	38·1	158	111	70·3
Gadag	,,	39	6	15·4	599	367	61·3
Dayalpur ...	1903	29	10	34·5	60	38	63·3
Punjab	1901-02	881	209	23·7	266,700	173,732	65·1
,, ...	1902-03	3,399	814	23·9	49,433	29,723	60·1
Aden	1905	37	14	37·8	368	294	79·9

[1] Compiled from Reports of the Plague Research Laboratory, Bombay, with the exception of the last three entries, which are extracted from Sticker's "Die Pest," part ii, p. 449.

tions issued by the Government of India. It will be seen that there is very little tendency for the death-rates of vaccinated and unvaccinated persons to vary together, a result which should make us regard the data either with suspicion or as being untrustworthy owing to insufficiency of observations.

We have now completed that part of our inquiry which relates to

the determination of the probability that inoculation does really confer some advantage on the inoculated person and we have seen that such appears to be the case both for typhoid and for cholera. It is therefore natural to inquire how we can measure the degree of advantage, and how we can compare the advantage derived in connexion with one disease with that obtained in the case of another malady. The discussion of these points compels one to face numerous difficulties and we shall have to examine certain matters not at first sight germane to the subject of the paper. This will form the topic of the next section.

Section III: The Meaning of Immunization and the Interpretation of Consistent Data relating to Two or more Epidemics.

In the previous sections we discussed in the first place the conditions necessary for insuring that the comparison of death- or attack-rates amongst immunized and non-immunized should be legitimate, and the methods for judging whether (supposing such comparability assured) any observed difference might, or might not, be due to the mere chances of sampling. We then proceeded to examine the available data for two diseases of present importance, typhoid and cholera, and to show how often the necessary conditions of comparability were unfulfilled or the numbers of observations insufficient to ensure complete certainty of judgment as to the significance of the results. In Section I we pointed out that on *any* hypothesis as to the meaning of immunization it seemed obvious that the death- or attack-rates amongst immunized and non-immunized should be very highly correlated, assuming the numbers of observations sufficient to render the errors of sampling relatively small; but have just indicated that this *a posteriori* test of consistence completely breaks down for more than one set of data. Suppose, however, that we can find some data for which this test of consistence is fulfilled and for which accordingly, within the margin of errors of sampling, there is a single-valued functional relation between the death- or attack-rates amongst immunized and non-immunized. We now wish to discuss the problem whether any hypothesis as to the meaning of immunization can be framed of sufficient definiteness to enable us to predict the form of that functional relation.

Let us consider first the assumptions implied in the method adopted by Professor Pearson for the treatment of such "fourfold tables" as have been given above—a method which has been very

largely used by Macdonell, Maynard, and others. One or two such tables were given by Professor Pearson in his original memoir on the method (1900, ii) and a table for deaths amongst vaccinated and unvaccinated small-pox patients is headed in the following form:—

TABLE XLIII.

Degree of effective vaccination	Strength to resist small-pox when incurred		Total
	Deaths	Recoveries	
Cicatrix absent	94	383	477
Cicatrix present	42	1,562	1,604
Total	136	1,945	2,081

It is assumed, as suggested by the headings, that we have really to do with two continuous variables: (1) Degree of effective vaccination, or perhaps it would be better to say immunity; those with "cicatrix absent" representing the patients of lower immunity, and those with "cicatrix present" the patients of higher immunity. (2) Strength to resist small-pox when incurred: those who die being the patients of lower, and those who recover the patients of higher, strength. The two groups, in each case, it is also assumed, are contiguous, but not overlapping, being sundered from each other at some one definite value of the continuous variable hypothesized, just as one might sunder short men from tall by taking all men as short who were under 5 ft. 4 in., and all as tall who were over that limit. Further, it is assumed that the distribution of the variables follow the normal law of error. The distribution of "degree of immunity" may accordingly be represented by the annexed diagram, fig. 3 (*a*), in which the proportion of vaccinated has been taken at 77·08 per cent., the value in the above table, and the distribution of "strength to resist small-pox when incurred" might be represented by a similar diagram. Finally, supposing the correlation to be also of the normal form, the coefficient of correlation between "degree of effective vaccination" and "strength to resist small-pox when incurred" can be calculated. Later writers who have dealt with similar tables have not always inserted general headings indicating what they assumed the supposed variables to be, but Macdonnell (1902 and 1903), for instance, terms the coefficient he obtains the correlation between "effectiveness of vaccination and strength to resist the disease" or between "vaccination and strength of resistance."

We should ourselves prefer to head the columns of the above table with some such words as "Mildness of case," and to substitute for

"Degree of effective vaccination" "Resistance of patient when infected." Whatever assumption is made as to the form of the distribution, the effect of vaccination is to increase the resistance of the patient, but the mildness or severity of the case is dependent not only on the patient's resistance but also on the quantity and quality of the infection, and also possibly on external circumstances, so that the correlation between resistance and mildness of case is not unity and the resistance of the patient measures only the *chance* of his recovery. Looking at the matter from this point of view, we may ask ourselves what is, on Pearson's hypothesis, the law relating "chance of recovery" to

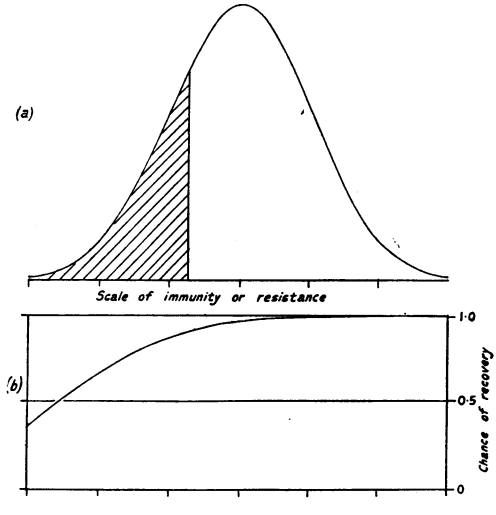

FIG. 3.

(*a*) Distribution of resistance on Pearson's theory; shaded area to the left being that of unvaccinated. (*b*) The chance of recovery for each degree of resistance.

"degree of effective vaccination," or, as we should prefer to term it, "resistance."

In fig. 4 let M_2M_2 be the vertical through the mean of the variable X_2, M the mean of the whole distribution, RR the line of regression through the means of rows, HH the vertical line of division, cutting off deaths to left and recoveries to right. Let r, σ_1, σ_2, be the correlation and standard deviations. Evidently as a horizontal array is moved downwards by, say, a unit the point at which HH cuts it moves to the left from the mean of the array by a constant amount,

that is, as the s.d. of the array is constant, by a constant fraction of the s.d. of the array. Hence the curve showing the chance of recovery for every value of the "resistance" X_1 must be a normal integral curve $C\,C$ as sketched on the right of the figure. To fix the curve we only require its centre (the point at which the chance is 0·5) and the distance from that centre of, say, the chances corresponding to normal deviations of unity—a distance which might well be termed the dispersion of the curve. The centre of the curve is given by the point P where the line $H\,H$ meets $R\,R$, for at that point $H\,H$ bisects the array, that is, the centre of the curve deviates from the mean of X_1 by an amount—

$$x_1 = \frac{1}{r}\,\frac{\sigma_1}{\sigma_2}\,h$$

where $h = H\,M_2$. To obtain the dispersion, note that at a distance s *from the point* P, $H\,H$ cuts the array at a distance $r\,\frac{\sigma_2}{\sigma_1}\,s$ from its mean; or, as the s.d. of the array is $\sigma_2\sqrt{1-r^2}$, at a fraction of the s.d. equal to $rs/\sigma_1\sqrt{1-r^2}$ from its mean. This fraction is unity—i.e., the array is cut by $H\,H$ at a distance from its mean equal to the s.d. when $s = \sigma_1\frac{\sqrt{1-r^2}}{r}$ The curve of chances is therefore now completely determined. For example, in Pearson's table above, the correlation on the assumption of normality is 0·5954, and $h/\sigma_2 = -1{\cdot}5114$. Hence $x_1 = -2{\cdot}5385\sigma_1$ and the dispersion $s = 1{\cdot}3495\ \sigma_1$. Fig. 3 (*b*) shows the curve in its proper position in relation to the scale of "degree of effective vaccination" or resistance, and the frequency distribution.

This form of law for the relation between resistance and chance of recovery looks to us a very reasonable one, but the assumption that the frequency distribution of resistance for vaccinated and unvaccinated together can be represented by a single normal curve as in fig. 3 (*a*) is quite another matter. The assumption implies that whatever the proportion of vaccinated (*a*) the resistance of no unvaccinated person is higher than that of a vaccinated person; (*b*) the resistance of the most resistant unvaccinated person is just equal to that of the least resistant vaccinated person[1]; (*c*) the frequencies of the most resistant unvaccinated persons and least resistant vaccinated

[1] Where the number of observations is large; where the observations are few, there may, of course, be appreciable differences between the resistances of adjacent individuals, ranked in order of resistance, but these differences decrease as the number of observations is increased and will be quite low except in the extreme tails of the curve.

persons are also equal. Surely an assumption with such implications is absurd on the face of it? Nor do matters look any better if we adhere to Professor Pearson's original wording: it is equally difficult to assent to an assumption which implies that no person without a cicatrix recorded is more "effectively vaccinated" than a person with cicatrix present; that the degree of effective vaccination of the most effectively vaccinated person with a cicatrix absent is always just equal to that of the least effectively vaccinated person with a cicatrix present; and that the frequency of the most effectively vaccinated persons with cicatrix absent is always just equal to the frequency of the least effectively vaccinated persons with cicatrix present. Possibly someone with greater imaginative powers than we

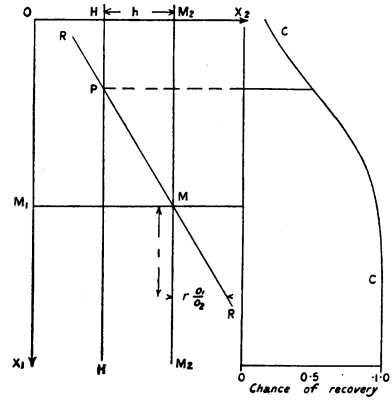

FIG. 4.

possess might be able to frame some hypothesis to illustrate how such a distribution might arise, even with the usual approximately random selection of persons for inoculation, but we must confess it has proved an impossible task to us. Matters become even more difficult when cases are considered in which varying proportions of the population have been inoculated under precisely the same conditions. Suppose, for example, that in one year one inoculates 25 per cent. of a certain population, choosing the persons for inoculation as near as may be at random: next year, with the same method and the same precautions one inoculates 75 per cent. What are the distributions of resistance, or "degree of effective vaccination"? Suppose the curve (*a*), fig. 5, to represent the distribution for the first year. How shall we place the curve for the second year? There seems no reason for supposing

that the bounding resistance between vaccinated and unvaccinated would be altered, so suppose we make these bounding values the same in the two years: curve (*b*) for the second year will then be placed as shown if we suppose the standard deviation also to be unaltered. But this makes the mean resistance of both vaccinated and unvaccinated much higher in the second year than in the first, in contradiction to our supposition that the whole process was carried out in precisely the

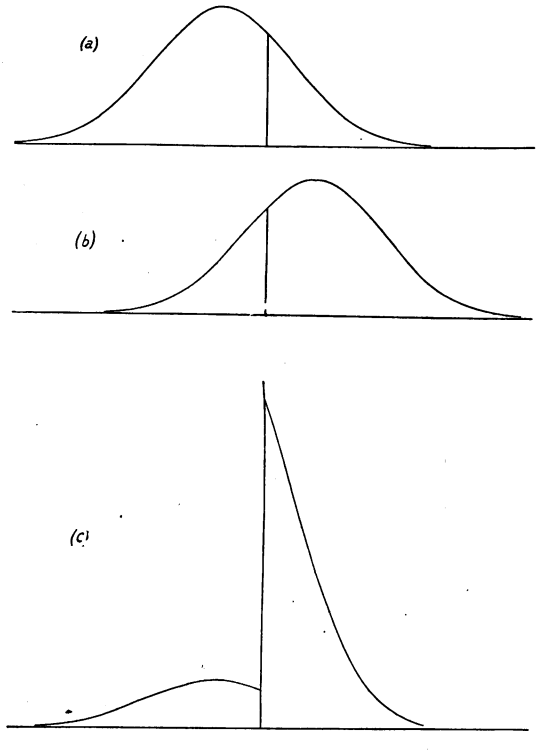

FIG. 5.

same way in both years, excepting only as regards the increase in the proportion of inoculated. If we decrease the standard deviation of the second curve, the agreement between the mean resistances in the two years becomes better for the inoculated and worse for the uninoculated: if we increase the standard deviation, the reverse is the case. If we give up the idea that the bounding resistance between inoculated and uninoculated is constant, we can make the mean resistance of the inoculated the same in the two years and also the mean resistance

of the uninoculated, *but by no shifting of the mean or alteration of the standard deviation can we make the distributions of the two groups similar in the first year and the second.* If in the first year the distribution of resistances is as given by fig. 5 (*a*) the distribution in the second year should, it seems to us, be given, not by a normal curve, but by the distribution (*c*) obtained from (*a*) by trebling all the ordinates for the vaccinated and taking one-third of each of the ordinates for the unvaccinated: the process being exactly the same in each year the form of the distribution of resistances for each group should be unaltered. We cannot see any way of avoiding this conclusion. The assumption that the distribution is of the form (*a*) in the one case is not consistent with its being of the same form in the second case.

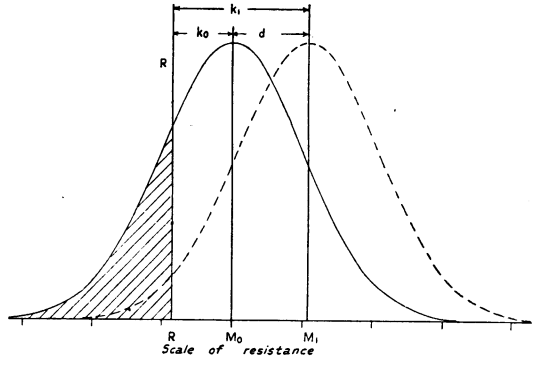

FIG. 6.

Distributions of resistance on Maynard's theory. Full line curve, unvaccinated; broken curve, vaccinated. M_0, M_1, means of unvaccinated and vaccinated. R, critical resistance, all of greater resistance recovering and of lower resistance dying.

The whole difficulty is, in fact, as one of us expressed it before (Yule, 1912, p. 638), that on this assumption one cannot effect a change in the numerical proportion of inoculated without changing them qualitatively at the same time.

This assumption as to the form of the distribution of "resistance" amongst the inoculated and uninoculated must in our judgment be rejected, and the question is whether some more reasonable assumption cannot be made. One has already been put forward by Maynard (1913, p. 67). "My conception of the problem," he writes, "is as follows: Members of a population possess the character 'power to

resist attack' in varying amounts, and the distribution of this population in regard to the character is continuous and probably capable of being represented, at least approximately, by a normal curve. Assuming inoculation to increase this 'power to resist attack,' its effect will be to bodily move the treated portion of this population in the direction of increasing amounts of this power, without, necessarily, altering the type of distribution." If, in fig. 6, the continuous normal curve represents the distribution of "resistance" amongst the uninoculated portion of the population in question, some curve more to the right, like the broken curve shown, will represent the distribution of resistance amongst that portion of the population that has been inoculated. We have shown the two curves as possessing the same standard deviation, as suggested by Maynard's wording, but this is not essential; the distribution for the inoculated population may possess a dispersion greater or less than that of the uninoculated group, the change in the dispersion being dependent on the way in which different individuals react to the dose. Further, the assumption that the two distributions are normal is evidently not essential, though necessary for simple mathematical treatment.

With this conception of Maynard's we find ourselves in entire agreement; it seems to us a simple and natural assumption which probably is a close representation of the facts. The two distributions are now shown, as they should be, as entirely independent; they may or may not largely overlap; the numbers in either group may be altered as largely as we please without altering the form of the distribution of resistance for that group.

A further assumption is also suggested by Maynard as to the portion of the population, whether inoculated or uninoculated, that is attacked or that dies as the case may be—the simplest assumption that can be made. Let us assume, that is to say, that all those are attacked or die who possess less than a certain degree of resistance, say the resistance R in fig. 6. Then the portions of the inoculated and uninoculated populations cut off to the left of the ordinate through R will be the portions attacked or dying, as the case may be. Some consideration suggests that this assumption is almost too simple to be a complete or true representation of what actually occurs, for it is difficult to suppose that in practice the chance of an individual being attacked depends only on his "resistance" and on no other circumstance. Let us take it, however, as a working hypothesis, develop its consequences and see how far it fits the facts.

As pointed out by Dr. Maynard, if the s.d.'s of the two curves are assumed to be the same, the results of a single epidemic suffice to determine the relative positions of the two curves. Thus, in the case of Professor Pearson's table which we have used as an illustration above, the proportion of recoveries amongst the unvaccinated ("cicatrix absent") is 0·8029. Reference to any table of the normal integral shows that this is equivalent to a normal deviation of 0·852 of the standard deviation. The mean M_0 of the distribution of resistance for the unvaccinated must therefore lie, taking the standard deviation as the unit of our scale, at a distance 0·852 of a unit to the right of the point we take as R. Similarly, the proportion of recoveries amongst the vaccinated is 0·9738, corresponding to a normal deviation 1·940, and M_1 the mean of the distribution for the vaccinated must lie at a distance 1·940 units to the right of R. The difference between the mean resistances of vaccinated and unvaccinated is $1{\cdot}940 - 0{\cdot}852$ or 1·088, the unit being the s.d. The diagram is drawn from these data, the numbers of vaccinated and unvaccinated being taken as equal.

But the assumption that the two standard deviations are equal seems to us unnecessary and undesirable. Suppose we take the s.d. of the distribution for the unvaccinated as unity, as before, the s.d. of the distribution for the vaccinated being σ_1, and the difference of the means $M_1 - M_0 = d$. A single epidemic no longer gives the necessary data for determining σ_1 and d. But suppose we have also data for a second epidemic, say of less intensity, cutting off all those whose resistance is less than S, then the distributions are determinate. For we have from fig. 6, noting that k_0 and k_1 are negative.

$$k_0 = k_1 + d \qquad (1)$$

The data give k_0 and the ratio k_1/σ_1, say a_1, so we may write this

$$k_0 = a_1\sigma_1 + d \qquad (2)$$

Inserting the values of k_0 and a_1 given by the two epidemics, we have a pair of linear equations for determining σ_1 and d.

But evidently, though we could determine in this way the two distributions of "resistance" consistent with the given data, no test would be afforded of the working hypothesis. To afford such a test we must have data for more than two epidemics and see how far they lead to the same constants for the distributions. The simplest way to carry out such a test is to work out the values of k_0 and a_1 for every epidemic in the series, plot them on a diagram and see whether, within the limits of fluctuations of sampling, the points obtained lie on a straight line.

The diagram will suffice to show whether there is an approximate fit, but for accurate work a straight line must be fitted to the observed points by some method that gives a precise answer, and further, we must be able to determine the probable true position of the more or less erroneous point in order to decide whether the differences between the observed and the true positions of the points are within the permissible limits. These problems are by no means easy to solve. If the probable errors of the two co-ordinates are the same, and if the distribution of the points may be taken as given by the normal law of error, the principal axis is the most probable true position of the line—it is the line which makes the sum of the squares of the perpendiculars from the observed points a minimum. But in fact the probable errors of the two co-ordinates are not the same, for the numbers of observations are never the same for inoculated and uninoculated: and, we may add, the probable errors are not the same for different points, as the number of observations varies from one epidemic to another. Further, there is little reason to suppose that the correlation is normal. In these circumstances it does not seem possible to give any thoroughly satisfactory general solution. Failing anything better, we decided to work with the principal axis; it gives a good, if not the theoretically best, solution—and a unique solution. It also may be shown that the true line must lie between the lines of regression, and this the principal axis does. We have not attempted any system of weighting the points to allow for the varying numbers of observations; this would have greatly increased the work, increased it indeed to a greater extent than the trustworthiness of the observations would seem to justify, and it did not seem worth while to attempt to develop such a method. The procedure we have used was therefore as follows :—

(1) Work out for each epidemic of the series the percentages of recoveries (or not attacked) amongst inoculated and uninoculated.

(2) Look up these proportions in a table of the normal integral and substitute the corresponding normal deviations, k_0 and a_1.

(3) Work out the means, standard deviations and correlation coefficient for the normal deviations.

If r, s_0, s_1, are the values found for the correlation and standard deviations, the angle θ which the principal axis makes with the axis of k_0 is, by a well-known formula, given by—

$$\tan 2\theta = \frac{2r\, s_0 s_1}{s_0^2 - s_1^2} \qquad (3)$$

The axis is then determined completely by the fact that it passes through the mean of the observations. To illustrate the work we have given the arithmetic for our first illustration in full.

But this is only the first part of the problem. Having obtained the line, or what we hope may be regarded as a fair approximation to its true position, what is the probable position on that line of the true point corresponding to a given observed point? The solution we have adopted is this: Suppose a perpendicular to be dropped from the observed point on to the principal axis; the foot of this perpendicular is approximately the true position of the point. Again, this solution is not strictly correct, but we judge it to be sufficiently so for our purpose. Here, as in obtaining the approximate position of the true line, we have endeavoured to minimize errors by selecting only cases in which the correlation is high. Supposing the equation obtained between k_0 and a_1—i.e., the equation to the principal axis—to be

$$k_0 = a_1\sigma_1 + d$$

the "corrected" value k_0' of any observed value k_0, that is the value which we judge to be approximately the true value, is given by

$$k_0' = \frac{a_1\sigma_1 + d + \sigma_1^2 k_0}{1 + \sigma_1^2} \qquad (4)$$

a_1 being the observed value associated with the observed k_0. The corrected value a_1' is most simply obtained by substituting k_0' in the equation to the line.

Illustrations.

After the remarks made at the end of the preceding section, it is not wonderful that we found considerable difficulty in discovering data which would afford any really satisfactory test of the theory. So inadequate in fact are the data, that we regard the cases given in the following pages as examples rather than as tests. For an adequate test we feel that recourse must be had to experiment, and on a somewhat large scale; failing such experimental data, we have in the meantime endeavoured to supplement our examples by considering cases in which the *differentia* between the two classes is no longer inoculation. It is evident, in fact, that the two curves of fig. 6 need not represent the distributions of resistance amongst inoculated and uninoculated respectively. The one might represent the distribution

of resistance, say, for males and the other for females: or the one for whites and the other for negroes. The theory, if applicable at all, should be applicable to many cases besides the comparison of inoculated and uninoculated individuals, indeed to cases in which we are not even concerned with data respecting sickness or mortality.

(*A*) *Cholera Inoculation.*—This was one of the first cases that we tried. In order that we might not have to deal with large errors of sampling and in the hope of thereby obtaining a more consistent series, we decided to exclude all cases in which χ^2 was less than 10. This would have left us with only three of Haffkine's cases, however, and this seemed rather few to deal with. But Nijland appears to have used the same method as Haffkine, so it seemed legitimate, as far as method of inoculation was concerned, to regard his material as on a par with that of Haffkine, though it differs in as far as his data refer in part to natives of the Dutch colonies and in part to Europeans. The Europeans we did not think it right to include with natives, and this left us with only three of Nijland's cases also. The six cases so sorted out afford the following data:—

TABLE XLIV.—CHOLERA INOCULATION: HAFFKINE'S AND NIJLAND'S DATA.

Place	Proportion not attacked		Normal deviations	
	Not inoculated	Inoculated	Not inoculated k_0	Inoculated a_1
(1) Calcutta ...	0·8776	0·9892	−1·1631	−2·2974
(2) Margherita ...	0·7687	0·9796	−0·7346	−2·0456
(3) Cachar ...	0·9698	0·9953	−1·8779	−2·5972
(4) Singaradja ...	0·9901	0·9988	−2·3301	−3·0370
(5) Semarang ...	0·9935	0·9990	−2·4838	−3·0902
(6) Selong ...	0·9941	0·9990	−2·5181	−3·0902

The two columns on the right show the normal deviations equivalent to the proportions of "not attacked" in the first two columns, and the diagram of fig. 7 shows how closely the points they give are grouped round a straight line.

These normal deviations give the following means, standard deviations, and coefficient of correlation, the subscript 0 referring to the not inoculated:—

$$M_0 = -1{\cdot}85127 \qquad M_1 = -2{\cdot}69293$$
$$s_0 = 0{\cdot}68255 \qquad s_1 = 0{\cdot}41205$$
$$r = 0{\cdot}99154$$

Hence from equation (3)—

$$\tan 2\theta = +1{\cdot}8836165$$
$$2\theta = 62^\circ\, 2'{\cdot}8$$
$$\theta = 31^\circ\, 1'{\cdot}4$$
$$\tan\theta = 0{\cdot}60141 \qquad \cot\theta = 1{\cdot}66275$$

As tan θ is the slope of the line to the axis of k_0, cot θ gives the ratio of a deviation in k_0 to a deviation in a_1, or the equation to the principal axis may be written—

$$k_0 + 1{\cdot}85127 = 1{\cdot}66275\ (a_1 + 2{\cdot}69293)$$

or—

$$k_0 = 1{\cdot}66275\ a_1 + 2{\cdot}62640 \qquad (5a)$$

This is the form in which we require the equation to compare with (2) on p. 149. It may also be convenient to have it for calculation in the form—

$$a_1 = 0{\cdot}60141\ k_0 - 1{\cdot}57956 \qquad (5b)$$

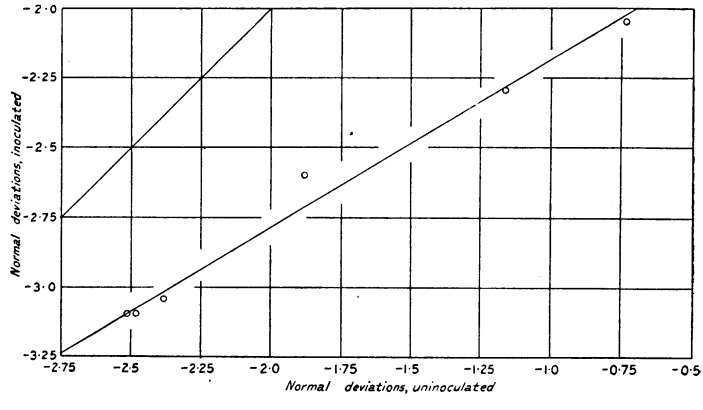

FIG. 7.

Cholera. Correlation between normal deviations for inoculated and uninoculated. The diagonal line in the upper left-hand corner is the line that would have been obtained if inoculation had been completely ineffective.

The line in fig. 7 is plotted from these equations. Equation (5a), comparing it with (2), tells us that, taking the standard deviation of the distribution of resistances for the uninoculated as unity, the standard deviation of the distribution for the inoculated is 1·66, and the mean of the same is 2·63 units higher than the mean resistance of the uninoculated. Fig. 8 shows the two curves drawn from these data, assuming the same number in each class, and the verticals numbered 1, 2, 3, 4, 5, 6 (determined by the method described below) correspond as closely as possible to the values of k_0 and a_1 in the last two columns of Table XLIV. In the Calcutta epidemic, that is to say, the proportions of inoculated and of uninoculated attacked are given very closely by the fractions of the two normal curves cut off to the left of the ordinate

through 1, all those with a resistance less than 1 being attacked. In the Margherita epidemic the proportions are similarly given by the fractions cut off to the left of the ordinate 2, and so on. While the mean resistance of the inoculated is very much higher than that of the uninoculated, their resistance is also much more variable; 94·3 per cent. of the inoculated have a higher resistance than the mean of the uninoculated, while only 4·3 per cent. of the uninoculated have a higher resistance than the mean of the inoculated.

We have stated above that the verticals, 1, 2, 3, 4, 5, 6, correspond "as closely as possible" to the observed values of k_0 and a_1 in Table XLIV. It is evident that they cannot, in general, correspond exactly, for equation (5) is not an exact, but only an approximate, relation

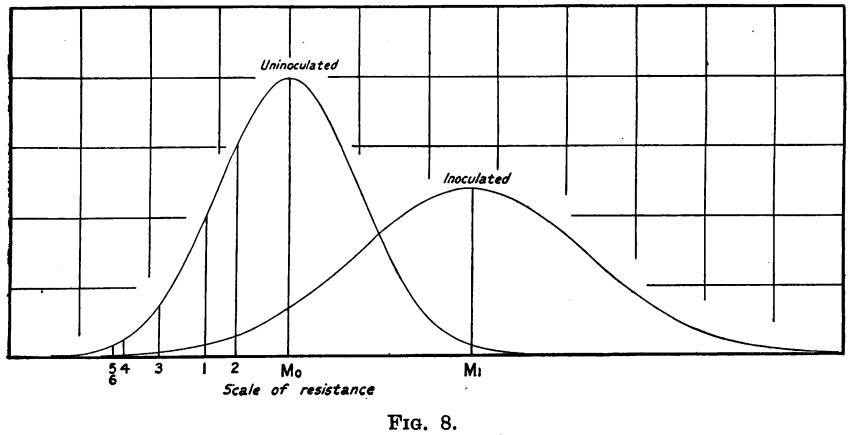

FIG. 8.

Cholera. Distributions of resistance of uninoculated (to the left) and inoculated (to the right). The side of one of the squares is the standard deviation for the uninoculated, and the two curves have equal areas—viz., ten squares. 1, 2, 3, &c., are the critical resistances for the six epidemics.

between k_0 and a_1, the closeness of fit being shown by the line in fig. 7. If, then, we plotted the verticals from the values of k_0, the values of a_1 would not be correct; if we plotted them from the values of a_1, the values of k_0 would not be correct, and positions intermediate between these two must be taken as probably the true positions. We estimated the true positions as described above, p. 151. Thus the corrected value of k_0 is given by

$$k_0' = \frac{1{\cdot}66275\, a_1 + 2{\cdot}76474\, k_0 + 2{\cdot}62640}{3{\cdot}76474} \qquad (6)$$

and for the Calcutta epidemic we have $k_0' = -1{\cdot}1712$ against $-1{\cdot}1631$ observed. The corrected value of a_1 is given by inserting the value of k_0' in the equation

$$a_1 = 0{\cdot}60141\, k_0 - 1{\cdot}57956 \qquad (7)$$

and we thus find $a_1' = -2{\cdot}2839$ (against $-2{\cdot}2974$ observed). Proceeding in this way we find the following corrected values—

TABLE XLV.

	k_0'	a_1'
(1)	−1·1712	−2·2839
(2)	−0·7453	−2·0278
(3)	−1·8285	−2·6792
(4)	−2·3549	−2·9958
(5)	−2·4912	−3·0778
(6)	−2·5164	−3·0929

A comparison of these values with those in Table XLIV shows that the corrections are not, in most cases, great, but a simpler comparison results, if we convert these normal deviations back into the corresponding proportions, as in the following table:—

TABLE XLVI.—CHOLERA INOCULATION: FIT OF THEORY AND OBSERVATION.

	Proportions not attacked			
	Observed	Calculated	Difference	Standard error	Ratio Diff./s.e.
(1) Not inoculated	0·8776	0·8790	−0·0014	0·01412	0·10
Inoculated	0·9892	0·9888	+0·0004	0·00619	0·06
(2) Not inoculated	0·7687	0·7720	−0·0033	0·03478	0·09
Inoculated	0·9796	0·9787	+0·0009	0·01010	0·09
(3) Not inoculated	0·9698	0·9662	+0·0036	0·00211	1·71
Inoculated	0·9953	0·9963	−0·0010	0·00090	1·11
(4) Not inoculated	0·9901	0·9907	−0·0006	0·00029	2·05
Inoculated	0·9988	0·9986	+0·0002	0·00054	0·37
(5) Not inoculated	0·9935	0·9936	−0·0001	0·00028	0·36
Inoculated	0·9990	0·9990	—	0·00035	0·00
(6) Not inoculated	0·9941	0·9941	—	0·00020	0·00
Inoculated	0·9990	0·9990	—	0·00045	0.00

The observed proportions are repeated in the first column; in the second are given the values corresponding to the "corrected" normal deviations above, and in the third column the differences between the corrected and observed values. These differences are then compared with their standard error. It will be seen that only a single difference exceeds twice the standard error, the twelve being grouped as follows: Less than the standard error, nine; exceeding the standard error, but not twice the standard error, two; exceeding twice the standard error, one. If the distribution of errors were normal the expectation would

be 8·2 : 3·3 : 0·5. The fit is thus a very fair one, but much stress must not be laid on this in view of the fact that our cases are stringently selected.

Fig. 9 shows the curve relating the proportions of not attacked amongst inoculated and uninoculated, corresponding to the straight line of fig. 7 relating the normal deviations: the diagonal across the figure is the form the curve would take if the proportions of not attacked amongst inoculated and uninoculated were equal. It may be remarked that, over the range covered by observation, the divergence from linearity of the relation between the two proportions is not conspicuous. Fig. 10 shows the form taken by the complete curve, from an intense

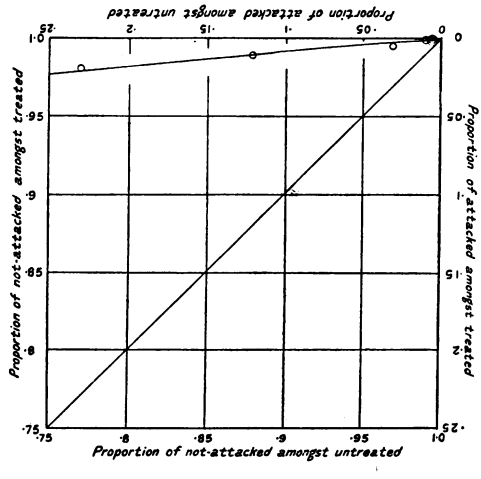

FIG. 9.

Cholera. Relation between proportions of not-attacked amongst treated and untreated; curve calculated from the two normal distributions. To read off the proportions of attacked reverse the diagram.

epidemic in which everyone is attacked, down to a very mild attack in which practically no one is attacked: observation only covers the last quarter of this curve. In an inset, we have drawn the extreme end of the curve, in the right-hand upper corner, to a greatly enlarged scale, in order to bring out a point regarding its form. In fig. 9 and in the general diagram of fig. 10 the curve appears to be tangential to the horizontal at its upper end, but it is really tangential to the vertical, as it crosses the diagonal at a value of p_1 equal to 0·999963 approximately and then turns upwards as shown in the inset. The reason for this is

obvious on considering the diagram of the two normal curves (fig. 8). The curve for the inoculated has a larger standard deviation than the other, and must therefore ultimately cut it towards the left and rise above it. We have, then, only to carry the vertical like 1, 2, 3, &c., sufficiently to the left to make the proportion of not attacked among the uninoculated greater than among the inoculated. The point at which this will occur is given by the two equations (5) (*a*) and (*b*) (p. 153) for this case, which give us $k_0 = a_1$ when

$$k_0 = 0{\cdot}60141\, k_0 - 1{\cdot}57956 = -3{\cdot}9628 \qquad (8)$$

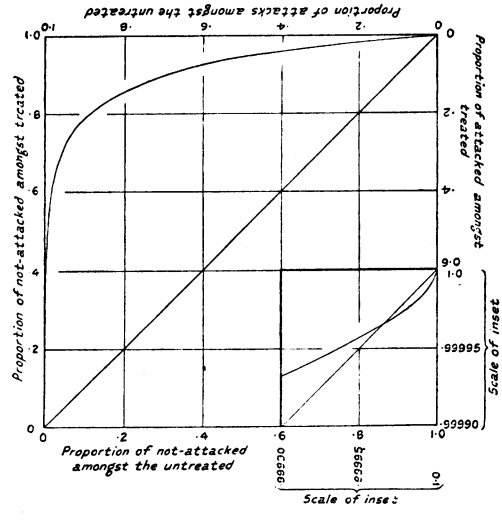

FIG. 10.

The complete curve of fig. 9. The right-hand end of the curve appears tangential to the horizontal. As a fact it is tangential to the vertical. The inset represents the right-hand extremity of the curve on a scale enlarged 4,000 diameters. When the proportion of not attacked amongst the untreated exceeds 0·999963 the proportion of not attacked is lower amongst the untreated.

that is, when the proportion of not attacked amongst the uninoculated is 0·999963 as stated above. It is extremely doubtful whether we could trust the normal law at deviations greatly exceeding three times the standard deviation, so that the point is, perhaps, of more theoretical than practical interest. But it is worth noting that distributions may exist in which the crossing point comes much lower in the scale of proportion not attacked. Thus, suppose the means of the two normal distributions of fig. 8 differed by only a single unit instead of by

2·626 units, then the proportions of recoveries would be equal when k_0 was 1·5088, or the proportion of not attacked 0·9343. When the proportion of not attacked among the uninoculated was greater than this, the proportion of not attacked among the inoculated would be less than the proportion among the uninoculated. It is an interesting question whether such distributions ever arise in practice; their existence would imply that the process of inoculation actually weakened resistance in some cases, so that, although the mean resistance was raised, the proportion of the population with less than a certain small resistance was increased and inoculation consequently became disadvantageous for mild epidemics, though advantageous for more or less virulent ones.

If the standard deviation of the resistance of the inoculated is less than that of the uninoculated, the reverse effect occurs, as the curves cut towards the right. Inoculation may then become disadvantageous for very virulent epidemics. Only if the standard deviations are the same the curves cannot cut more than once, and inoculation, if advantageous for an epidemic of one intensity, is advantageous for all other intensities.

(*B*) *Swine Fever.*—Table XLVII shows the results of serum treatment for swine fever in Hungary, according to information obtained

TABLE XLVII.—SWINE FEVER. RESULTS OF INOCULATION WITH SERUM ALONE OF HEALTHY SWINE. HUNGARY.

(From a pamphlet entitled "Serum Treatment for Swine Fever," issued by the National Federation of Meat Traders' Associations, in September, 1913.)

Number of herds	Treated	Died		Untreated	Died	
		Head	Per cent.		Head	Per cent.
15	1,419	—	—	899	97	10·9
15	2,183	54	2·4	1,572	201	12·7
3	360	29	7·9	219	57	26·0
6	1,108	179	16·1	1,029	461	44·8
3	190	40	21·0	254	120	47·2
4	229	70	30·4	1,035	615	59·4
46	5,489	372	8·8	5,008	1,551	30·9
6	599	383	63·8	659	382	57·9
52	6,088	755	12·4	5,667	1,933	34·1

through the British Consulate. It is not quite clear on the face of it how the table is compiled. We assume that it is to be read as follows. There were fifteen test herds in which none of the treated

swine died; of the untreated swine in these herds, of which there were 899, 97, or 10·9 per cent., died. There were also fifteen herds in which some, but (probably, we think) not more than 5 per cent. of the treated swine died, 54 out of 2,183 in these herds dying, or 2·4 per cent.; of the untreated swine in these herds, numbering 1,572, 201, or 12·7 per cent., died—and so on. But we do not understand why the group of six herds in which 63·8 per cent. of the treated swine died is tabulated separately at the foot of the table, nor why the results for this group are so divergent. Omitting the first line of the table, inasmuch as a percentage zero, which is presumably due only to the chances of sampling, corresponds to a normal deviation of infinity, and also omitting the last divergent group of six herds, we have the following data :—

TABLE XLVIII.—SWINE FEVER. DATA USED AND DERIVED FROM TABLE XLVII.

Proportions of survivals			Normal deviations		
Untreated		Treated	Untreated (k_0)		Treated (a_1)
0·8721	...	0·9753	−1·1364	...	−1·9651
0·7397	...	0·9194	−0·6424	...	−1·4011
0·5520	...	0·8384	−0·1307	...	−0·9879
0·5276	...	0·7895	−0·0692	...	−0·8047
0·4058	...	0·6943	+0·2384	...	−0·5081

The diagram of fig. 11 shows how closely the points given by these normal deviations lie to a straight line. Working out the means, standard deviations and coefficient of correlation, we find :—

$$M_0 = -0{\cdot}34806 \qquad M_1 = -1{\cdot}13338$$
$$s_0 = 0{\cdot}48520 \qquad s_1 = 0{\cdot}50667$$
$$r = 0{\cdot}99610$$

whence

$$\tan 2\theta = -22{\cdot}99343$$
$$2\theta = 92^\circ\ 29'{\cdot}4$$
$$\theta = 46^\circ\ 14'{\cdot}7$$
$$\tan\theta = 1{\cdot}0444 \qquad \cot\theta = 0{\cdot}9575$$

The equation to the principal axis may therefore be written in either of the forms—

$$\left.\begin{aligned} k_0 &= 0{\cdot}9575\, a_1 + 0{\cdot}7371 \quad \text{(a)} \\ a_1 &= 1{\cdot}0444\, k_0 - 0{\cdot}7699 \quad \text{(b)} \end{aligned}\right\} \qquad (9)$$

The line of fig. 11 is drawn from these equations and fig. 12 shows the two distributions of resistance, for equal numbers of inoculated

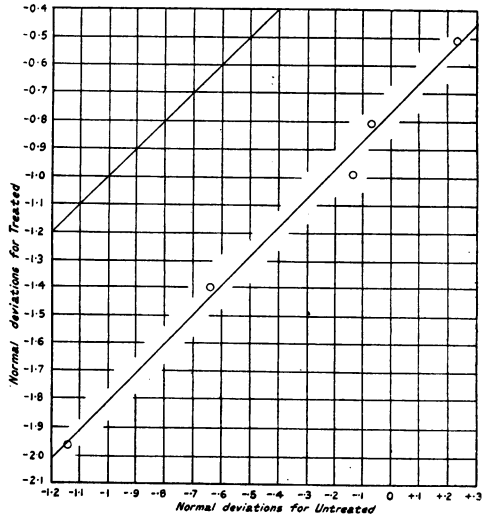

FIG. 11.

Swine fever. Relation between normal deviations corresponding to proportions of recoveries amongst treated and untreated animals respectively.

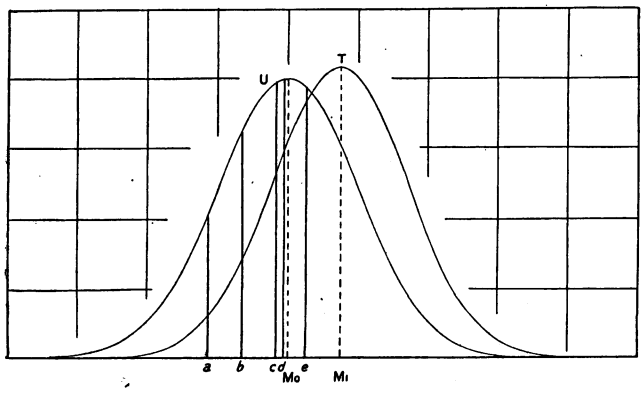

FIG. 12.

Swine fever. Distributions of resistance of untreated animals (to the left) and treated (to the right). The areas of the two curves are equal; if each represents 1,000 animals, one square is 100. The unit of the horizontal scale is the standard deviation of the untreated. *a*, *b*, *c*, *d*, *e*, are the critical resistances for the five groups of herds affected by outbreaks of varying severity.

and uninoculated. Taking the standard deviation for the uninoculated as unity as before, the standard deviation for the inoculated is 0·9575, and the mean of the inoculated is 0·7371 unit higher than the mean of the uninoculated. We are uncertain whether the swine fever data are fatality-rates in the strict sense or deaths in terms of a population some members of which were never attacked; the cholera data are, of course, incidence-rates. But the conclusions as regards the distributions of resistance are very different. In the case of cholera the means differed by more than 2·5 units, here they differ by less than 0·75. For cholera, the dispersion of the resistance of the inoculated was much greater than that of the uninoculated. For swine fever the dispersion of the resistance of the inoculated animals is slightly the lower of the two. A glance at the two diagrams shows the nature of the difference more effectively than any verbal description.

Evaluating the "corrected" normal deviations as in the last illustration, we find—

TABLE XLIX.

	k_0'	a'_1
(1)	−1·1406	−1·9611
(2)	−0·6226	−1·4201
(3)	−0·1714	−0·9489
(4)	−0·0505	−0·8226
(5)	+0·2448	−0·5142

and hence we have the following comparison of the observed proportions of survivals with the calculated proportions:—

TABLE L.—SWINE FEVER.—FIT OF THEORY AND OBSERVATION.

		Proportions of recoveries		Difference	Standard error	Ratio Diff./s.e.
		Observed	Calculated			
(1)	Untreated	0·8721	0·8730	−0·0009	0·00842	0·11
	Treated	0·9753	0·9751	+0·0002	0·00332	0·06
(2)	Untreated	0·7397	0·7332	+0·0065	0·0297	0·22
	Treated	0·9194	0·9222	−0·0028	0·0144	0·19
(3)	Untreated	0·5520	0·5680	−0·0160	0·0155	1·03
	Treated	0·8384	0·8287	+0·0097	0·0111	0·87
(4)	Untreated	0·5276	0·5201	+0·0075	0·0313	0·24
	Treated	0·7895	0·7946	−0·0051	0·0296	0·17
(5)	Untreated	0·4058	0·4033	+0·0025	0·0153	0·16
	Treated	0·6943	0·6964	−0·0021	0·0304	0·07

The fit in this case is even closer than in the last, as the difference in only one case exceeds the standard error. Fig. 13 shows graphically the fit of the theoretical curve to the observations, and the form of

the entire theoretical curve. At the lower left-hand corner the curve actually sweeps round below the diagonal and terminates tangentially to the horizontal, this being an example of the theoretical case in which inoculation may become disadvantageous for very virulent epidemics. But the point at which the diagonal is cut by the curve lies even farther beyond the sphere of practical politics than in the last illustration. From equations (9) we find that $k_0 = a_1$ when

$$k_0 = 1{\cdot}0444\, k_0 - 0{\cdot}7699 = 17{\cdot}339$$

and this value of k_0, if the normal law continued to hold good for such extreme deviations, would correspond very nearly to a proportion

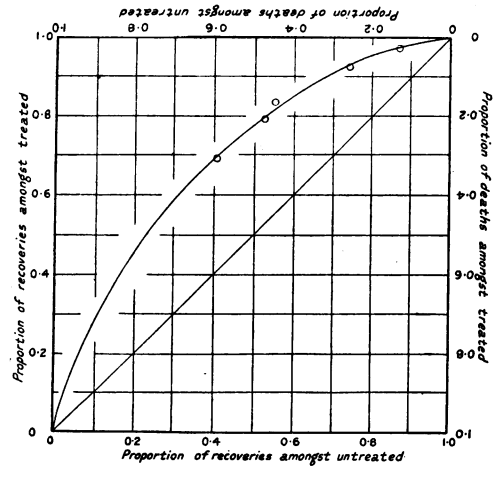

FIG. 13.

Swine fever. Relation between proportions of recoveries or deaths in the case of treated and untreated animals. The observed points are given by small circles; the curve is calculated from the two frequency distributions of resistances given in fig. 12.

of survivals represented by a unit in the sixty-sixth place of decimals.

The two preceding illustrations may seem very inadequate as examples of the theory with which we are dealing, but they are the only illustrations based on statistics of inoculation that we can give, for the reasons already mentioned. We turned, therefore, to data from other sources, but even here found it difficult to discover thoroughly satisfactory material. Unless the number of observations

is large, fluctuations of sampling are of considerable importance, the correlation between the death-rates or the equivalent normal deviations in the two groups is only moderate, and it becomes doubtful how far the principal axis represents the true position of the line we are endeavouring to determine; it is also difficult to find material in which the death-rates exhibit any very great range, and unless the range is considerable a very poor test is afforded of the linearity of the regression between the normal deviations.

The first illustration that follows is on the whole the best of several similar cases that we tried.

(*C*) *Diphtheria Fatality at Two Age-groups.* — From the Annual Reports of the County Medical Officer for London we extracted the figures for notified cases and deaths from diphtheria at the age-groups 0-5 and 5-10, in each of the years 1892 to 1913, and worked out the fatality-rates for each sex. Calculating the equivalent normal deviations, we then took each sex separately and, regarding the age-group 0-5 as the standard, or as if it were the group of the uninoculated, proceeded as before to find the distributions of resistance. The results were as follows, the subscript 0 referring to the lower age-groups:—

Males.

M_0 −0·77566 M_1 −1·31126
s_0 0·37420 s_1 0·26252
r 0·98010
$\tan\theta$ 0·69673 $\cot\theta$ 1·43528
$k_0 = 1{\cdot}43528\, a_1 + 1{\cdot}10637$
$a_1 = 0{\cdot}69673\, k_0 - 0{\cdot}77084$ (10)

Females.

M_0 −0·75261 M_1 −1·28105
s_0 0·36156 s_1 0·29197
r 0·97849
$\tan\theta$ 0·80384 $\cot\theta$ 1·24402
$k_0 = 1{\cdot}24402\, a_1 + 0{\cdot}84104$
$a_1 = 0{\cdot}80384\, k_0 - 0{\cdot}67607$ (11)

Figs. 14 and 15 show the distribution of the points about the principal axes for the respective cases, and figs. 16 and 17 the distributions of resistance. It will be noticed that the correlations are lower in these cases than in the last two illustrations, but the numbers of observations are not large, ranging from about 1,000 to 2,700. We have not worked out the complete comparison of the proportions derived from the corrected normal deviations with the actual proportions,

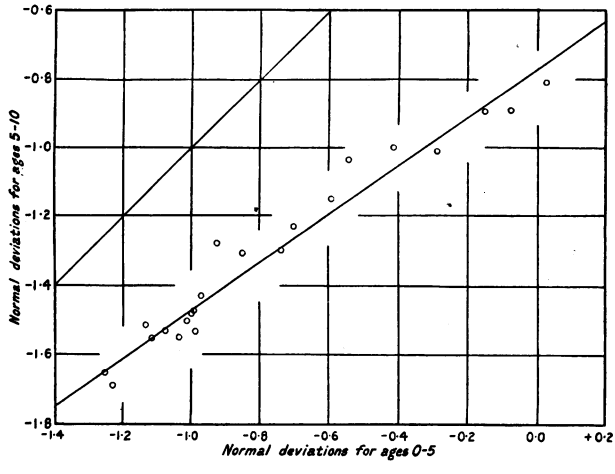

FIG. 14.

Diphtheria, males. Correlation between normal deviations at ages 0-5 and 5-10.

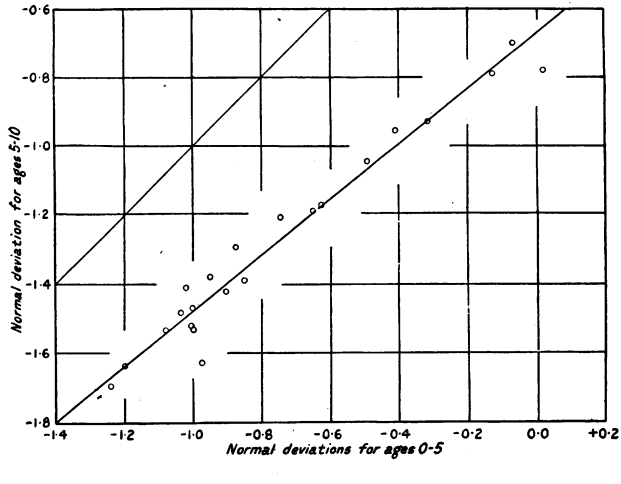

FIG. 15.

Diphtheria, females. Correlation between normal deviations at ages 0-5 and 5-10.

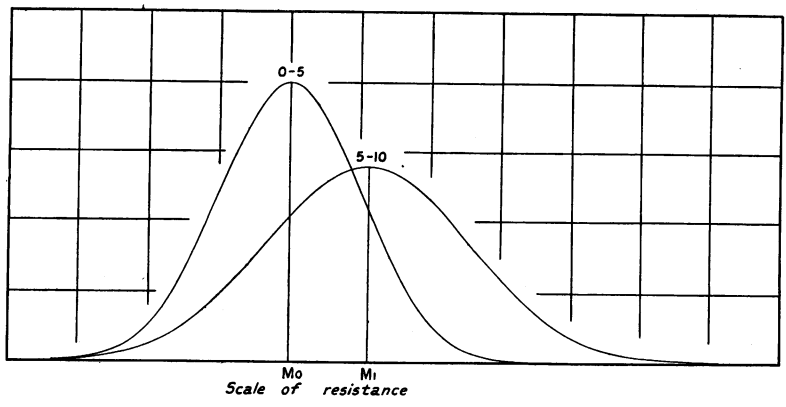

FIG. 16.

Diphtheria, males. Curves of resistance at ages 0-5 and 5-10. The area of each curve is ten squares, and the side of a square is the s.d. of the distribution for 0-5.

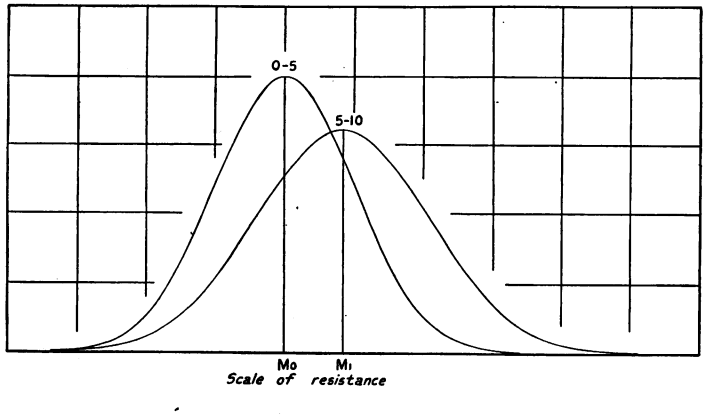

FIG. 17.

Diphtheria, females. Curves of resistance at ages 0-5 and 5-10. The area of each curve is ten squares, and the side of a square is the s.d. of the distribution for 0-5.

and of the differences with their probable errors, as in illustrations *A* and *B*, but the following test on one of the points does not suggest that deviations exceed the limits of sampling. One of the most divergent points is that for females in 1905, for which the co-ordinates are

$$k_0 - 0{\cdot}97421 \qquad a_1 - 1{\cdot}62841$$

The numbers of observations being 1,285 and 1,160 respectively. We find the following observed and calculated proportions of recoveries.

	Proportion of recoveries 0-5		5-10
Observed	0·8350	...	0·9483
Calculated	0.8547	...	0·9364
Difference	0·0197	...	0·0119
Standard error	0·0104	...	0·0065
Ratio diff./standard error	1·89	...	1·83

In each case the difference is less than twice the standard error.

So far we have treated each sex by itself, but it was natural to attempt to carry the work rather further and to endeavour to obtain all four distributions on a common scale. Females 0-5 showed slightly the highest mean fatality, judging from the mean normal deviation, so it was decided to take them as the standard. But it is clear that a difficulty may arise. Correlating the normal deviations for males 5-10 with the normal deviations for females 0-5, we determine the distribution of resistances for the former directly in terms of the distribution of resistances for the latter. But correlating males 5-10 with males 0-5, and males 0-5 with females 0-5, we have another indirect determination of the distribution for the first in terms of the distribution for the last, and the two determinations may not agree.

We find the following equations for the distributions of males in terms of females 0-5 :—

Males 0-5 against Females 0-5 (r = 0·99062)

$$k_0 = 0{\cdot}96591\ a_1 - 0{\cdot}00339 \qquad (12)$$

Males 5-10 against Females 0-5 (r = 0·99597)

$$k_0 = 1{\cdot}37900\ a_1 + 1{\cdot}05562 \qquad (13)$$

From equations (10) and (12) we have

s.d. Males 5-10 / s.d. Males 0-5 = 1·43528
s.d. Males 0-5 / s.d. Females 0-5 = 0·96591

Therefore

s.d. Males 5-10 / s.d. Females 0-5 = 1·43528 × 0·96591 = 1·38635

as compared with the directly determined value from equation (13) 1·37900.

The agreement seems to us very satisfactory and the mean of the two results, 1·383, may be taken as close to the truth. As regards the means :—

By (10)—

Mean Males 5-10 − mean Males 0-5 = 1·10637 × s.d. Males 0-5

by (12)—

Mean Males 0-5 − mean Females 0-5 = −0·00339 × s.d. Females 0-5

also by (12)—

s.d. Males 0-5 = 0·96591 × s.d. Females 0-5.

Therefore, finally,

Mean Males 5-10 − mean Females 0-5 = 1·06526 × s.d. Females 0-5

as compared with the directly determined value 1·05562.

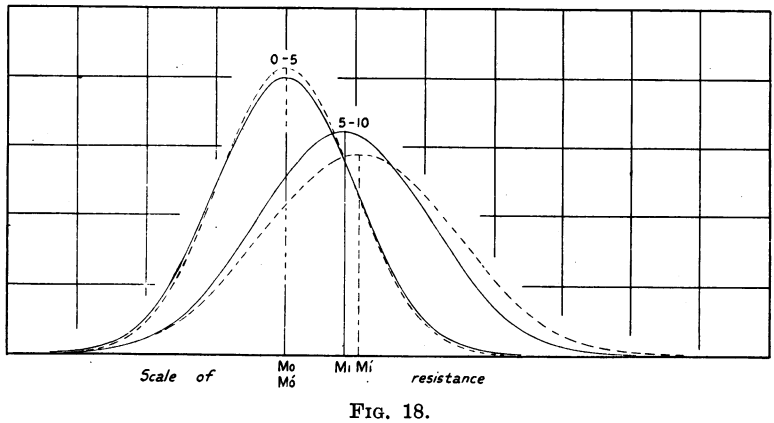

FIG. 18.

Diphtheria, males and females. Curves of resistance of the two sexes at ages 0-5 and 5-10 on the same scale. Full line curves, females; broken curves, males. The area of each curve is ten squares, and the side of a square is the s.d. of the distribution for females 0-5.

Again the agreement seems very fair, and we may take the average value, 1·060, as representing approximately the truth.

The distributions of resistance for the four groups therefore stand approximately as follows :—

	Mean	Standard deviation
Males 0-5	−0·003 ...	0·966
Females 0-5, taken as	0 ...	1·000
Females 5-10	0·841 ...	1·244
Males 5-10	1·060 ...	1·383

The distributions are shown in fig. 18. The mean percentages of recoveries for the four groups in the order given—that is to say, the

percentages corresponding to the mean normal deviations—are 0·781, 0·774, 0·900, 0·905, so that, as stated above, females aged 0-5 have a lower average proportion of recoveries than males of the same age, *but they show a higher (or sensibly the same) mean resistance.* The advantage obtained by males, within the range of fatalities in our data, is due to the lower variability of their resistance. For $k_0 = -0{\cdot}1$ almost exactly, or a proportion of recoveries 0·5398, males and females show the same fatality: for lower proportions of recoveries amongst females that sex has the advantage, if our assumption holds good. The most striking feature of the four distributions is the rise in the standard deviation *pari passu* with the rise in the mean resistance.

The next example we do not regard as a very satisfactory one, as the correlations are relatively low. We only give the results in brief, as they are interesting to compare with the foregoing.

(*D*) *Scarlet Fever Fatality at Two Age-groups.*—The data were precisely parallel to those of illustration *C*—viz., fatality-rates for London at the two age-groups 0-5 and 5-10, but at the moment when the work was done we only had at hand data for the years 1892-1910 and these were consequently the years utilized. The distributions of resistance were first determined for each of the two sexes separately, and then for the two sexes together, taking females aged 0-5 as the standard as in the last case. The results for the two sexes separately were as follows:—

Males.

M_0 −1·4827 M_1 −2·0757
s_0 0·1403 s_1 0·1148
r 0·8332

$$\left.\begin{aligned} k_0 &= 1{\cdot}2715\, a_1 + 1{\cdot}1567 \\ a_1 &= 0{\cdot}7865\, k_0 - 0{\cdot}9097 \end{aligned}\right\} \qquad (14)$$

Females.

M_0 −1·5327 M_1 −2·1228
s_0 0·1503 s_1 0·1084
r 0 8198

$$\left.\begin{aligned} k_0 &= 1{\cdot}4847\, a_1 + 1{\cdot}6190 \\ a_1 &= 0{\cdot}6735\, k_0 - 1{\cdot}0905 \end{aligned}\right\} \qquad (15)$$

In this case again we only tested the magnitude of the divergences of the points from the line in a single case. The point that looked to us one of the most divergent was that given by the observations for males in 1900. For this point the co-ordinates are

k_0 −1·5040 a_1 −2·2801

and the numbers of observations are 2,052 and 2,300 respectively. We find the following observed and calculated proportions of recoveries:—

	Proportion of recoveries	
	0-5	5-10
Observed	0·9337	0·9887
Calculated	0·9447	0·9848
Difference	0·0110	0·0039
Standard error	0·00549	0·00220
Ratio diff./s.e.	2·00	1·77

In the first case the difference is just over twice the standard error, in the second case less than twice. It does not seem, therefore, that on the whole the divergences exceed the limits of errors of sampling.

From the additional calculations necessary for reducing all distributions to females aged 0-5 as standard we find—

Males 0-5 *against Females* 0-5.

r 0·9726.

$$\left.\begin{aligned} k_0 &= 1{\cdot}0736\, a_1 + 0{\cdot}0591 \\ a_1 &= 0{\cdot}9314\, k_0 - 0{\cdot}0551 \end{aligned}\right\} \quad (16)$$

Males 5-10 *against Females* 0-5.

r 0·8211.

$$\left.\begin{aligned} k_0 &= 1{\cdot}3861\, a_1 + 1{\cdot}3444 \\ a_1 &= 0{\cdot}7215\, k_0 - 0{\cdot}9699 \end{aligned}\right\} \quad (17)$$

The indirect calculation gives—

s.d. Males 5-10 / s.d. Females 0-5 = 1·3651.

against the directly calculated value 1·3861, mean 1·3756. The indirect calculation gives the difference of the means as 1·3009 times the s.d. for females 0-5 against the directly determined value 1·3444 (a poor agreement this time), mean 1·3226. The distributions of resistance for the four groups (fig. 19) therefore stand approximately as follows :—

	Mean	Standard deviation
Females 0-5, taken as	0	1·000
Males 0-5	0·059	1·074
Males 5-10	1·323	1·376
Females 5-10	1·619	1·435

We note the same characteristic as in the last illustration of increasing standard deviation accompanying the rising mean. Further, while the mean normal deviation for the females aged 0-5 corresponds to a slightly higher proportion of recoveries than the mean normal deviation of the males of the same age—viz., 0·9373 against 0·9309—the mean resistance of the former is lower than that of the latter. This is just the reverse of what we noticed in the case of the diphtheria fatalities, where a *lower* proportion of recoveries amongst the females was accompanied by a *higher* resistance. In the present case again the

females exhibit the highest and lowest mean resistances, while in the case of diphtheria the males took the extreme positions.

Let us add that neither in this case nor the last do we desire to put forward such conclusions as can be drawn from the forms found for the respective distributions of resistance, as anything more than interesting illustrations of the points that may be suggested by the use of the method on adequate data. The data used, in our opinion, are not adequate to enable anything but the most tentative conclusions to be drawn, and we have not yet discussed several difficulties connected with the method.

(*E*) *Infant Mortality for the Two Sexes.*—It occurred to us that this might form an interesting case as, by taking districts of very varying

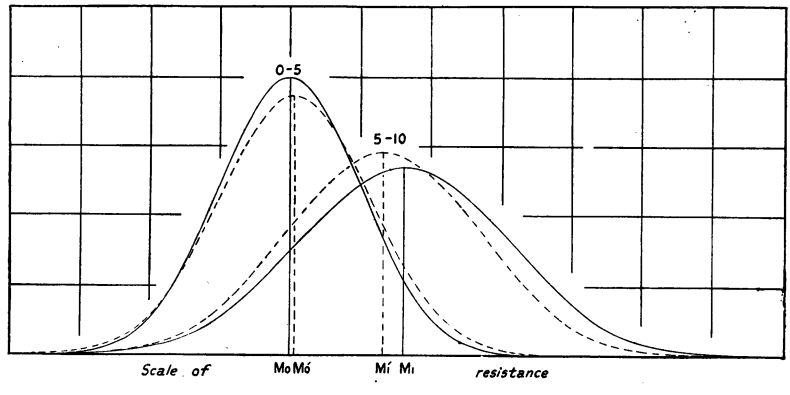

FIG. 19.

Scarlet fever, males and females. Distributions of resistance of the two sexes at ages 0-5 and 5-10 on the same scale. Full line curves, females; broken curves, males. The area of each curve is ten squares, and the side of a square is the s.d. of the distribution for females 0-5.

character, or possibly different countries, data for a considerable range of mortality could be obtained. Ultimately we used two sets of data:—

(I) We first turned to the English registration data given in the decennial supplements for 1881-90 and 1891-1900. As we wanted large districts, so as to keep fluctuations of sampling small, we first of all picked out certain urban registration districts, and then, in order to extend the range of mortality, added some counties of a more or less rural character. The first set of districts was thus made up as follows: (*a*) thirty London Registration Districts, 1881-90 and 1891-1900; thirty Lancashire Registration Districts, 1881-90 and 1891-1900; thirteen

Rural Registration Counties, 1881-90 and 1891-1900; total for the two decades, 146. (*b*) We then added, in order further to extend the material in the direction of low mortality, forty Administrative Counties in 1912—namely, those counties in which the number of births for each sex exceeded 1,500—making the total number of observations used, 186. (*c*) Finally, in order a little further to extend the line in the direction of high mortality, we added data for thirteen Bavarian districts in 1911, making the total number of observations used 199.

(II) The second set of data was still more heterogeneous. In the two volumes published under the title of "Statistique Internationale" by the Statistique de la France and edited by M. March, data are given for most of the principal countries of the world showing infantile mortality for each sex separately by quinquennial periods, from the earliest epoch at which the statistics could be obtained (cf. the first volume, pp. 463-64, and the second volume, p. 124). Plotting the mortality for the females against that for the males, it seemed that a few countries gave rather divergent results, showing a female mortality that was relatively high compared with the male mortality.[1] For this reason the figures for Sweden before 1800, Italy, Bulgaria, Servia, and Roumania were omitted. The data included, which, so far as one could judge from the rough graphic test, seemed fairly homogeneous, covered the following numbers of quinquennial periods in the countries stated:—

England and Wales	12
Scotland	11
Ireland	9
Denmark	12
Norway	14
Sweden	22
Finland	9
Russia	3
Austria	11
Hungary	3
Switzerland	8
Prussia	7
Bavaria	14
Saxony	11
Wurtemberg	8
Baden	8
Holland	12
Belgium	13
France	21
New South Wales	2
New Zealand	2
Total	212

[1] A few obvious errors in the rates were corrected by reference to the original data, also given in the volume.

These sets of data gave the following results :—

I (*a*) The subscript 0 refers to the male sex.

$$\begin{array}{ll} M_0 \; -0{\cdot}98029 & M_1 \; -1{\cdot}10294 \\ s_0 \quad 0{\cdot}13509 & s_1 \quad 0{\cdot}13629 \end{array}$$
$$r \; 0{\cdot}98108$$
$$\left.\begin{array}{l} k_0 = 0\;9910\, a_1 + 0{\cdot}1127 \\ a_1 = 1{\cdot}0091\, k_0 - 0{\cdot}1138 \end{array}\right\} \qquad (18)$$

I (*b*)

$$\begin{array}{ll} M_0 \; -1{\cdot}05515 & M_1 \; -1{\cdot}17975 \\ s_0 \quad 0{\cdot}19261 & s_1 \quad 0{\cdot}19538 \end{array}$$
$$r \; 0{\cdot}98610$$
$$\left.\begin{array}{l} k_0 = 0{\cdot}9858\, a_1 + 0{\cdot}1079 \\ a_1 = 1{\cdot}0144\, k_0 - 0{\cdot}1094 \end{array}\right\} \qquad (19)$$

I (*c*)

$$\begin{array}{ll} M_0 \; -1{\cdot}03512 & M_1 \; -1{\cdot}16008 \\ s_0 \quad 0{\cdot}20490 & s_1 \quad 0{\cdot}20652 \end{array}$$
$$r \; 0{\cdot}98675$$
$$\left.\begin{array}{l} k_0 = 0{\cdot}9920\, a_1 + 0{\cdot}1157 \\ a_1 = 1{\cdot}0080\, k_0 - 0{\cdot}1167 \end{array}\right\} \qquad (20)$$

II

$$\begin{array}{ll} M_0 \; -0{\cdot}91042 & M_1 \; -1{\cdot}02427 \\ s_0 \quad 0{\cdot}23977 & s_1 \quad 0{\cdot}23086 \end{array}$$
$$r \; 0{\cdot}99880$$
$$\left.\begin{array}{l} k_0 = 1{\cdot}0386\, a_1 + 0{\cdot}1534 \\ a_1 = 0{\cdot}9628\, k_0 - 0{\cdot}1477 \end{array}\right\} \qquad (21)$$

In the case of the original 146 districts of set I (*a*) the three worst points were given by the Strand District in 1891-1900, the City of London in the same decade, and Garstang in 1881-90. Using the same method as before, we find the following values for the observed and calculated death-rates and the ratio of the differences to their standard errors.

District	Observed	Calculated	Diff./s.e.
Strand	0·280	0·267	1·41
	0·215	0·229	1·62
City	0·252	0·243	1·09
	0·197	0·207	1·29
Garstang	0·113	0·119	0·76
	0·091	0·084	1·04

These figures do not suggest that in set I (*a*) at all events the deviations observed are greater than may be expected from fluctuations of sampling alone. Fig. 20 illustrates the correlation for the complete set I (*c*), and fig. 21 the correlation for set II.

It will be noticed that the results of the first series are fairly self-consistent, the additions to set I (*a*) of the forty additional districts included in I (*b*) and the thirteen additional districts included in I (*c*) not making any great alteration in the values obtained for the standard deviation and mean of the female distribution of resistance, which centre round 0·99 and 0·11 respectively. This set therefore shows the females to be slightly more resistant than the males and also to have a slightly lower variability, by which the advantage is increased.

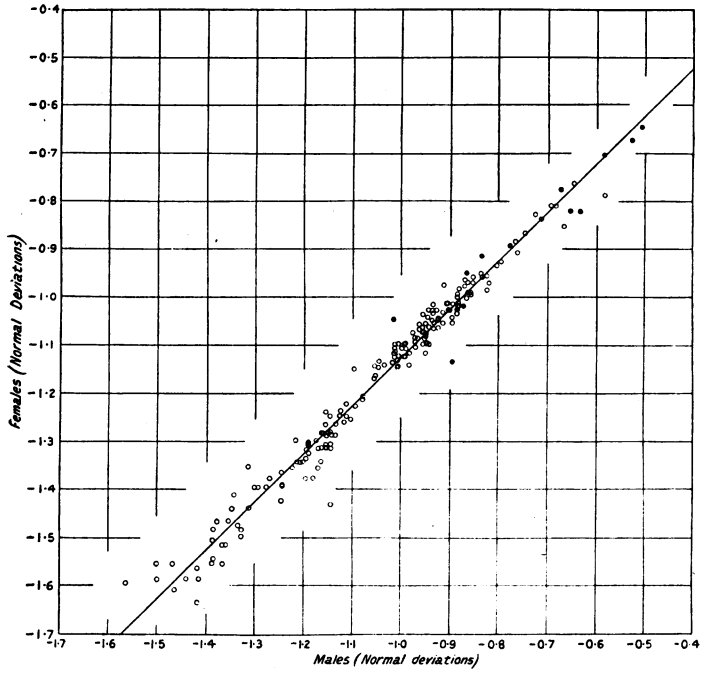

FIG. 20.

Correlation between normal deviations of infantile mortality, male and female. Circles, English districts; blocked-in circles, Bavarian districts.

But the second set of data gives a rather different result. This shows the mean resistance of the females as relatively higher, 0·15 units greater than the male mean instead of 0·11, but their standard deviation is higher than that of the males instead of lower, being 1·04 units approximately, not 0·99. It might appear that for possible moderate rates of mortality this greater variability might more than counterbalance the greater mean resistance of the females; but this

is not so. By equations (21) the distributions shown in set II would only give equal mortalities for males and females when

$$k_0 = 0{\cdot}9628\, k_0 - 0{\cdot}1477 = -3{\cdot}970$$

and this value of k_0 corresponds to the impossibly low mortality of 36 per million.

The correlation given by the second set of data is the higher of

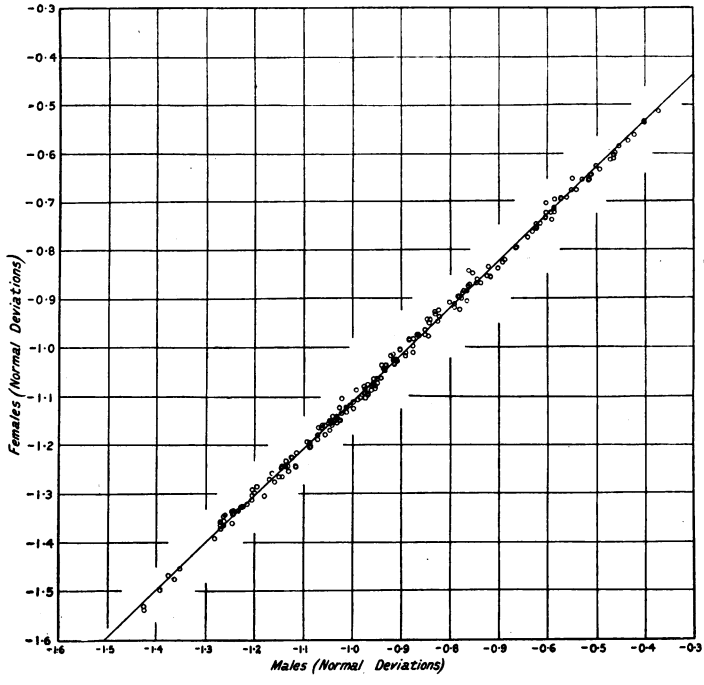

FIG. 21.

Correlation between normal deviations of infantile mortality, male and female, in various countries. Data from the *Statistique Internationale*.

the two, and the result given by that set is the less likely to be considerably affected by fluctuations of sampling. The question therefore arises whether the s.d. of the resistance of the females is really and generally higher than that of the males as given by the second set, the fact that the first set gives a lower value being perhaps only due to a fluctuation of sampling. Unfortunately we cannot give a definite answer to this question as, so far as we know, the standard error

of the slope of the principal axis ($\tan \theta$) has not yet been evaluated by anyone else and we have not evaluated it ourselves. Probably it is of the same order as the slope of the regression lines, but it may, we think, be higher. If we take it as of the same order, we may regard it as given approximately by $\sqrt{1 - r^2}/\sqrt{n}$ (the geometric mean of the standard errors of the two regressions). This comes to 0·0115 for set I (*c*) and only 0·00336 for set II, or 0·0120 for the difference between the two. The actual difference is 1·0386 − 0·9920 or 0·0466, which is 3·9 times the standard error so estimated. On the whole it seems, therefore, unless we have considerably underestimated the standard error, that the difference is greater than is likely to be due to errors of sampling alone.

When we had reached in our work the stage to which we have at present carried the reader, we must confess that we began to feel somewhat uncomfortable concerning the accuracy with which the theory appeared to hold good; it held good more closely than we felt it had any right to.

In the first place, it did not seem to us very likely *a priori* that the distributions of resistance would be normal distributions. It seemed more probable, as there must be something like an absolute limit towards low or zero resistance, that both distributions would be skew, with longer tails towards high than towards low resistance and that the distribution with the higher mean would be less skew than the other. Why then, we asked ourselves, did the regression generally seem to be very closely linear? To investigate this question, we thought it would be interesting to work out the actual curves that would result if we assumed the two distributions of resistance to be skew distributions of known form. If we took this form to be that of one of Professor Pearson's well-known skew frequency curves, the calculation of the area up to each point on the scale would necessitate a good deal of work, so to make a first rough test we assumed the number of observations within successive intervals of the scale to be given by a binomial series. The following were then taken as pairs of distributions of resistance:—

I (*a*) The binomial $p = 0{\cdot}2$, $q = 0{\cdot}8$, $n = 20$. (*b*) The same binomial placed one interval higher on the scale.

Both distributions in this case were therefore very skew and equally skew.

II (*a*) The same binomial as the above. (*b*) The binomial $p = 0{\cdot}3$,

$q = 0{\cdot}7$, $n = 20$. These binomials had the same origin, so that the mean of the first is 4, and of the second 6.

III (*a*) The same binomial as in (*a*) above. (*b*) A normal curve, placing the mean of the curve at the upper limit of the interval on the binomial which includes the frequency of six successes, and making the interval 0·5 of the s.d. of the normal curve. These distributions give the following correspondent normal deviations:—

Binomial $(0{\cdot}2 + 0{\cdot}8)^{20}$	The same one place higher	Binomial $(0{\cdot}3 + 0{\cdot}7)^{20}$	Normal curve
−2·273	—	−3·156	−3·0
−1·483	−2·273	−2·428	−2·5
−0·820	−1·483	−1·807	−2·0
−0·224	−0·820	−1·243	−1·5
+0·331	−0·224	−0·715	−1·0
+0·857	+0·331	−0·211	−0·5
+1·361	+0·857	+0·274	0
+1·849	+1·361	+0·746	+0·5
+2·326	+1·849	+1·209	+1·0
+2·794	+2·326	+1·665	+1·5
+3·239	+2·794	+2·115	+2·0

The diagram (fig. 22) shows the lines plotted from the three pairs of distributions mentioned. It was plotted by us, of course, to a considerably larger scale and we confess that the result surprised us. The curvature of the line is in each case very slight—extremely slight in Case II, which represents most closely the forms we thought such distributions likely to assume—and only becomes evident owing to the considerable range covered. But the range covered by the data in most of our illustrations is very small, amounting to a little more than a unit of the scale, as will be seen from the following summary:—

Illustration	Normal deviations of untreated or standard		Range
	Lowest	Highest	
Cholera	−2·52	−1·16	1·36
Swine fever	−1·14	+0·24	1·38
Diphtheria (males)	−1·26	+0·02	1·28
,, (females)	−1·24	+0·02	1·26
Scarlet fever (males)	−1·70	−1·27	0·43
,, ,, (temales)	−1·81	−1·27	0·54
Infantile mortality (II)	−1·42	−0·37	1·05

Considering the smallness of the range it is not surprising that curvature of the degree shown in the figure does not make itself evident, and it is clear that we must not regard the apparent linearity of the

regression over the very small range that is in general available as proving that the assumption of normality is justified. Given a pair of skew distributions of resistance with certain means and standard deviations, there may well be a pair of " equivalent " normal distributions which will give, over a wider range than will generally be available in practice, so nearly the same relation between normal deviations that the two will hardly be distinguishable within the limits of sampling; but these normal distributions will not have the same means and standard deviations as the original distributions. Thus the line for the two binomials of Case II is fitted very closely over the whole range by the

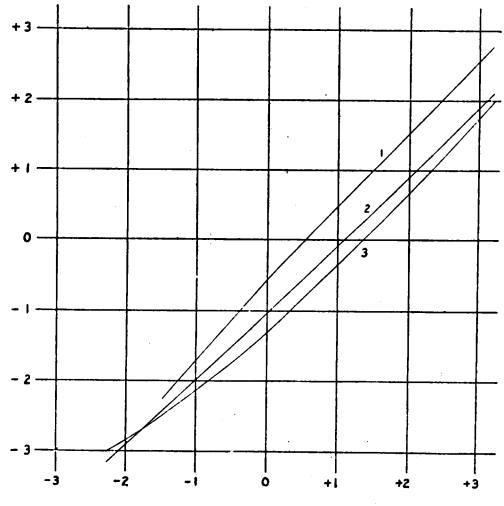

FIG. 22.

Relation between normal deviations, assuming distributions of resistance to be (1) the binomial $(0{\cdot}2 + 0{\cdot}8)^{20}$, and the same moved one place up; (2) the same and the binomial $(0{\cdot}3 + 0{\cdot}7)^{20}$; (3) the binomial of the first case and a normal curve.

straight line, of which the constants were determined by our usual method,

$$x = 1{\cdot}0456\, y + 1{\cdot}0609$$

That is, the ratio of the standard deviation of the second distribution to that of the first is given as 1·0456 and the difference of the means as 1·0609 times the s.d. of the first. But, actually, the standard deviations are 1·78885 and 2·04939, ratio 1·14565: while the means

are 4 and 6, difference 2, and 2/1·78885 is 1·11804. The binomial distributions have, therefore, more widely different standard deviations and means than the two normal curves which are, from our present standpoint, approximately equivalent.

The test applied is, of course, a rough one, and we quite admit that the subject deserves much more investigation than we have at present been able to give to it, but we think it may be definitely concluded that such illustrations as have been given above must not be held to prove, can hardly even be said to create any strong presumption, that the distributions of resistance, assuming the hypothesis otherwise true, are of the normal form given by the calculations. The solution of the problem as to the forms of the distributions which is given by our method is formal rather than true, and our difficulty falls to the ground.

But our second difficulty was of a more important character, and was brought strongly into prominence by the last numerical illustrations given. The hypothesis assumes that the mortality (supposing that we are dealing with the alternatives death or recovery) is strictly selective in the most stringent sense of the term: the disease cuts down all those with a resistance less than the critical value R, and permits all those with a resistance higher than R to survive. Surely, even in the case of fatality-rates this is rather a strong assumption to make? It seems at first sight easy to say: All the patients have the same disease—what determines the patient's death or survival? Why, of course, the patient's resistance. But may not the quantity and quality of the infection enter into the decision as well? If so, the patient's resistance determines, not the definite fact of death or recovery, but only the *chance* of death. The case against so simple a hypothesis as that on which we have been working is the same if we are dealing not with fatality-rates, but with attack-rates. Even if the population has been limited to that in which all the individuals are more or less at risk, are they all equally exposed to infection in the same degree? If not, again, the *chance* of infection may be a single-valued function of the resistance (of course, a resistance different from the resistance to death), but infection cannot be determined by the unique fact of the resistance being greater or less than some critical value R. In our last illustrations we were dealing not even with the fact of infection by a single definite disease, but with the general death-rate from all sorts of diseases. How can it be said in such a case that death or survival is uniquely determined by the quality of the patient? It must be admitted

that the element of chance or circumstance enters largely. How then does it come about that the hypothesis fits the facts as closely as it does?

The solution, we believe, may be found by considering how, on our present hypothesis, Professor Pearson's theory of the fourfold table must be transformed. Referring to Table XLIII on p. 142 and fig. 4, and following Professor Pearson with a slight change of nomenclature, we regarded OX_1 on the figure as the scale of "resistance," and OX_2 as the scale of "mildness of attack." The vertical HH cuts off deaths to left and recoveries to the right, and a horizontal line, not inserted on the figure, would, on Professor Pearson's theory, cut off vaccinated below and unvaccinated above. Now, consider the same table from the standpoint of the present method. We have been regarding the distributions of vaccinated and unvaccinated as distinct normal distributions and have usually spoken of them as distributions of *resistance*. But we may equally well regard them, not as distributions of resistance on the axis OX_1 of fig. 4, but as the distributions of "mildness of attack" on the axis OX_2, *mildness of attack being only correlated with resistance*, and the distributions being shifted to the left, across HH, without altering their forms or their relative positions, if the intensity of the epidemic is increased. The correlation surface for "resistance" and "mildness of attack" will now be a compound normal surface, though two simplifying assumptions may well, we think, be made. We may assume, that is to say, (1) that the line of regression RR is continuous through both surfaces, otherwise there would be a discontinuity in the relation between resistance and mean mildness of attack, and (2) that the standard deviation of the array is the same for both surfaces, otherwise there would be a discontinuity in the law relating chance of recovery to resistance. On these assumptions, the law relating chance of recovery to resistance is the same as on Pearson's hypothesis—viz., the curve of chances is a normal integral curve.

We suggest then that we have determined, by the method used in our illustration, the two curves corresponding to "mildness of attack" in this case of small-pox fatality, and not the two curves of resistance. In many cases it may be difficult to give a definite interpretation or a precise name to these two curves; we should be hard put to it, for example, to assign a term to the scale if we were dealing with a fourfold table for vaccination and attack, or for sex and infantile mortality. But the point is this: It may be either (*a*) that the epidemic

or other cause of death (or disease) simply cuts down all those with a resistance lower than a certain critical value, or (*b*) that chance enters largely into the matter, and we can only say that the chance of death is a function of the resistance; if this function be of a certain form (the normal integral function if the correlation be normal) then (*a*) is necessarily a possible geometrical interpretation of the facts even though (*b*) is the true interpretation. For the given form of function relating chance of recovery to resistance, which seems a reasonable sort of form, we cannot distinguish between the two hypotheses (*a*) and (*b*)—so long, of course, as we cannot obtain some independent measure of resistance. The process may not be so crude, nor so strictly selective as it appears on our diagrams, like fig. 8 or fig. 12, showing a couple of normal distributions of "resistance" and a series of verticals marking off the quick from the dead in epidemics of increasing intensity. The relative forms of the two distributions of resistance may in truth be more or less different, and the division between the quick and the dead not sharp but blurred; given not by a vertical which moves across the two distributions, if the intensity of the epidemic is raised, and separates the survivors at increasing values of the resistance, but given by a curve of chances of the form of the normal integral curve which is moved to the right without altering its form, so that every chance of death p, which formerly corresponded to a resistance x, now corresponds to, say, a resistance $x + d$.

The reader will naturally ask if we cannot go further than this and determine in some way the true distributions and the form of the curve of chances. The answer, so far as we can see, must be in the negative. Certainly the fourfold table tells one absolutely nothing. The table (XLIII) on p. 142, for example, merely informs us that there are 1,604 observations in the correlation surface between resistance and mildness of attack for the vaccinated, and 477 in the corresponding surface for the unvaccinated. The vertical HH of fig. 4 cuts off forty-two observations to the left in the first surface and ninety-four in the second. All this tells us nothing about the correlation in either surface or the forms of the total distributions. Supposing from a *series* of consistent epidemics we have determined the forms of the "mildness of attack" or pseudo-resistance curves, we are no better off. To take an imaginary numerical illustration: Suppose these pseudo-resistance curves have unit standard deviations and that the distance between their means is also unity. We know nothing about the correlation. Assume that the standard deviation of the array (row) is 0·5. We know nothing about the s.d.'s

of the true resistance curves either; let us make them unity also. Then from the value of the s.d. of the array we have $1 - r^2 = 0{\cdot}25$, $r = 0{\cdot}8660$, and the second condition gives the slope of the regression line to the vertical as r. Hence the distance between the means of the true resistance distributions is $1/0{\cdot}8660$ or $1{\cdot}155$. The point round which the chance-of-recovery curve centres depends on the severity of the epidemic; suppose that k_0 in the pseudo-resistance curves is -1, then the chance-of-recovery curve centres round the deviation $-1/r$ or $1{\cdot}155$. The dispersion of the chance-of-recovery curve (cf. above, p. 144) is the s.d. of the array divided by r, or $0{\cdot}577$. Fig. 23 illustrates the case supposed. The two normal distributions drawn with unbroken

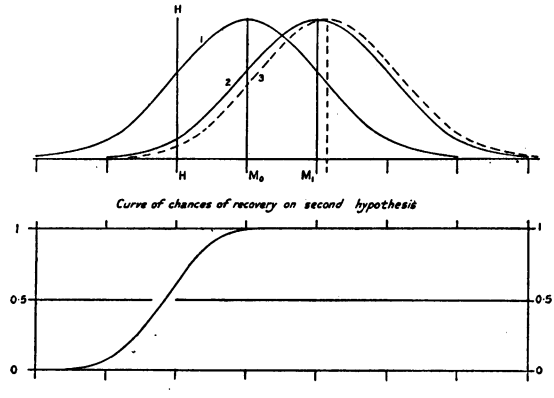

FIG. 23.

lines are the distributions arrived at by the method used in our illustrations. Retaining the same distribution as that assumed for the unvaccinated, the broken curve represents the true distribution for the vaccinated, the curve of chances of recovery being shown in the lower part of the figure. If the critical resistance for the first pair of distributions is moved by a unit, the centre of the chance of recovery curve must be moved by $1{\cdot}155$ units in the same direction. Precisely the same law will then be arrived at for relating the death-rates among vaccinated and unvaccinated.

It may, perhaps, be felt by the reader that the arguments of the preceding paragraphs have so mangled the original hypothesis that little vitality is left in it. We have shown, he may say, that the normal distributions which we have calculated may not represent the truth, for

the distributions may, in fact, be skew; and we have then proceeded to suggest that the calculation of the death-rates amongst immunized and unimmunized persons, by regarding the epidemic as eliminating all who possess less than a certain critical resistance, is a mere piece of mathematical jugglery. But such a pessimistic opinion we would by no means endorse.

We think it is in all probability essentially true that the resistances of immunized and non-immunized persons vary *inter se* and that these distributions may largely overlap. That, it seems to us, is the vital part of the hypothesis. We are, unfortunately, not in a position to make any direct determination of these distributions: they may very likely be asymmetric. It is an advantage, and not a disadvantage, if we can arrive at very much the same conclusions as regards the relative death-rates amongst the two classes on the supposition that they are normal.

Further, we do think it is probably true that the elimination of all individuals with less than a certain resistance cannot be regarded as representing what, in fact, happens during an epidemic, or by deaths in almost any other circumstances: chance must enter into the matter more or less. No one would be better pleased than ourselves if some way could be found for measuring resistance directly and determining the chance of death for persons of each grade of resistance. In the meantime it seems to us a distinct advantage that the element of chance may possibly be ignored and that we may be able to determine distributions of "pseudo-resistance" such that the elimination of individuals with pseudo-resistances less than R', $R' + d'$, $R' + 2d'$, $R' + 3d'$, &c., gives the same death-rates amongst the two classes compared, as would the application of the unknown law of chances of death to the true distributions of resistance, the law remaining unaltered in form but centring at values of the true resistance, R, $R + d$, $R + 2d$; and so on.

Section IV: The Measurement of the Relative Efficiencies of Different Immunization Processes.

We now come to the two questions which are of interest to the practical man—viz., granted that in certain cases, say cholera and typhoid, the death- or incidence-rates upon the uninoculated are higher than the corresponding values for the inoculated, and that these divergences cannot be dismissed as mere chance events, in which case

did the process of immunization produce the better result? If the typhoid results are good enough to entitle us to enforce inoculation in the case of troops moving into a typhoid-infected area, are the cholera results sufficiently good to authorize our taking the same step in the case of troops likely to be exposed to infection?

Professor Pearson answered such questions as these by referring to the values taken by his normal coefficient in the respective cases. Thus in his original memoir (1900) he obtained the value of 0·5954 for the correlation between what he termed "strength to resist small-pox when incurred" and "degree of effective vaccination." He then under the title "effectiveness of antitoxin treatment" prepared three tables for diphtheria, the horizontal dichotomy in each case being with "antitoxin, 1896" and "without antitoxin, 1894." The vertical dichotomies were (*a*) "recoveries" and "deaths"; (*b*) "requiring tracheotomy," and "not requiring it"; (*c*) "recoveries" and "deaths." The first two tables refer to laryngeal cases only: the third includes the cases of all children under five years of age. The normal coefficients in the three cases were 0·4708, 0·2385, 0·2451. Professor Pearson comments on these results as follows: "The three coefficients are all sensible as compared with their probable errors, and that between the administration of antitoxin and recovery in laryngeal cases is substantial. But the relationship is by no means so great as in the case of vaccination, and if its magnitude justifies the use of antitoxin, even when balanced against other ills which may follow in its train, it does not justify the sweeping statements of its effectiveness which I have heard made by medical friends. It seems until wider statistics are forthcoming a case for cautiously feeling the way forward rather than for hasty generalization" (op. cit. p. 45). We should conclude from this passage that Professor Pearson would rank the effectiveness of immunization processes in the order of the normal coefficients calculated from the available statistics, assuming, of course, that the differences are statistically significant.

We fear, however, that the questions propounded above cannot be answered in this simple fashion: a justification of this statement will be afforded by a closer examination of the problem. At the very outset one is faced with difficulties of definition. Thus, Professor Pearson speaks of the "effectiveness" of an immunization process. But this word may carry more senses than one. In what precise, numerical sense are we to understand it? To make the point clear, imagine the case of two diseases the fatalities of which were invariable, being

in every epidemic 50 and 5 per cent. respectively in the case of unimmunized persons. Now, suppose that in each case a process of immunization was employed which resulted in the fatality-rates of inoculated persons being reduced to 30 per cent. and zero respectively. In a certain sense the effectiveness of the immunization process in the second case is perfect, for the fatality-rate cannot be lower than zero. But, in another sense, the effectiveness is greater in the previous case because there the fatality-rate is reduced by 20 per cent. instead of 5 per cent. only. Evidently both these points of view are of importance. From the standpoint of the practical sanitarian a reduction of mortality from 50 to 30 per cent. is a greater achievement than a diminution from 5 per cent. to nothing. On the other hand, looking at the matter from the point of view of the student of immunizing processes, it might well be argued that the reduction to zero testifies to the establishment of a complete degree of immunity, the fact that the starting point was only 5 per cent. limiting the public advantage to be gained from the establishment of such a condition, but not its scientific interest. These considerations immediately suggest the desirability of employing two terms, in clearly defined senses.

We propose to define the *advantage* of an immunization process as the difference between the fatality-rates (or incidence-rates) of the unimmunized and immunized populations. The *efficiency* we propose to define as the ratio of the advantage to the fatality-rate (or incidence-rate) amongst the unimmunized—i.e., as the ratio of the numbers who are saved by the process to the numbers who might be saved. The practical application of these ideas is not, however, a simple matter. In our imaginary illustration we assumed that the fatality-rates of the two classes were invariable from epidemic to epidemic. But this is never the case, with the result that both measures vary from epidemic to epidemic of the same disease and that the figures for epidemics of two different diseases are not comparable. We may illustrate the first point on the cholera data. Assuming the two distributions of resistance of fig. 8, we have—

Incidence-rate	Advantage	Efficiency
0·9	0·6906	0·7673
0·8	0·6585	0·8231
0·7	0·5969	0·8527
0·6	0·5232	0·8720
0·5	0·4429	0·8858
0·4	0·3584	0·8960
0·3	0·2709	0·9030
0·2	0·1815	0·9075
0·1	0·0906	0·9060

It will be evident from this example that the "advantage" and the "efficiency" of any process of immunization are functions not only of the nature of the process, but also of the incidence-rate or fatality-rate. It hardly seems necessary to enforce the point by algebraic analysis. It follows that a comparison of the "advantages" or "efficiencies," when our data consist only of statistics relating to a single epidemic of each of two different diseases, may be seriously misleading. Were we in possession of a long series of epidemics in each case, the ranges of incidence- or fatality-rates being considerable, average values might reasonably be contrasted. Suppose, for example, fig. 24 represents the curve of p_2's for all values of p_1 (as in figs. 10 and 13), then RS is p_1, QR is $p_2 - p_1$ and PR is $1 - p_1$. The ratio of the area $AQCR$ to $ABCR$

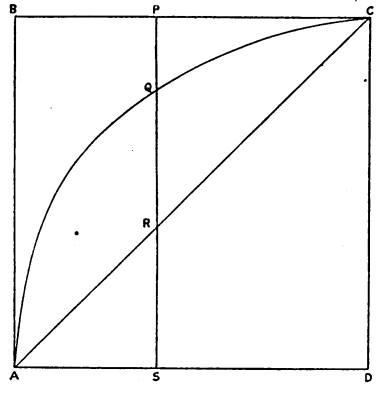

FIG. 24.

represents in a certain sense the mean efficiency, over the whole possible range of incidence- or fatality-rates. The area of the triangle ABC being 0·5, twice the area $AQCR$ is the mean efficiency. The area $AQCR$ itself is the mean advantage—i.e., the mean efficiency in this sense is twice the mean advantage. Applying this method to the case of swine fever and cholera (figs. 10 and 13), by a process of quadrature the mean efficiencies prove to be 0·404 and 0·799 respectively. These figures are not really comparable, since in one case we are dealing with deaths and in the other with attacks, but the example shows the sort of comparison that might have been made had the graphs referred to homologous data.

The practical limitations of the method are, however, apparent; not only is the actually observed range of incidence or fatality usually

very narrow, but we have assumed that the hypothesis discussed in the previous section is rigidly true and consequently permits of accurate extrapolation. Even when the data of several epidemics are available, the comparison of the efficiencies of two processes of immunization is attended with doubt and difficulty. When single epidemics are in question, the process becomes impossible. All we have then got is a single point on each of the two loci defining the inter-relations of the incidence- or fatality-rates of immunized and non-immunized persons.

If the soundness of the preceding arguments be admitted, it necessarily follows that no coefficient of correlation or association will furnish us with answers to the questions proposed at the beginning of this section. We believe that tentative answers can only be given on the basis of several epidemics, and with the help of some such process as that above described. In view, however, of the popularity enjoyed by certain of these coefficients and the fact that they have frequently been used in such investigations, it may be of interest to pass them in review.

(*A*) *Professor Pearson's Normal Coefficient.* — We have briefly described the nature of this coefficient, which has hitherto been by far the most frequently used, in the last section (pp. 141, 142). The fourfold table is assumed to be of the form—

	Variable X_1		
Variable X_2	$< H$	$> H$	Total
$< K$ ...	a ...	d ...	$a + d$
$> K$ ...	b ...	c ...	$b + c$
Total ...	$a + b$ ...	$c + d$ ...	N

and the coefficient of correlation is calculated on the assumption that the correlation between X_1 and X_2 is normal, the distribution being divided sharply into the four compartments by planes parallel to the axes of measurement. Now this coefficient might have been at once excluded, as one of us has previously pointed out (Yule, 1912, pp. 587-588), on the ground that it is not a function of p_1 and p_2 alone, but also of the proportion of vaccinated. For given values of p_1 and p_2 it is a maximum when the proportion of vaccinated is 50 per cent. Thus if the fatality-rate is 10 per cent. amongst the vaccinated and 40 per cent. amongst the unvaccinated, the normal coefficient is 0·57 when the proportion of vaccinated is 50 per cent., 0·50 when it is 15 per cent., 0·43 when it is 5 per cent., and only 0·35 when it is 1 per cent. If the proportion of

vaccinated be as high as 85 per cent. the coefficient falls from the maximum value to 0·53. What practical meaning can be assigned to a coefficient that behaves in such a way?

Further, if the arguments of the last section be admitted, the whole foundation of the method is cut away. For if the "vaccinated" and "unvaccinated" do not correspond to the mere dichotomy of one continuous distribution of resistance, but there is one distribution for the vaccinated and another for the unvaccinated, the two frequencies a and d constitute one frequency surface, the two frequencies b and c constitute another. There are no data for determining the correlation in either surface alone or in the two together. The correlation is, as we have stated above (p. 180), completely indeterminate.

It seems to us that in this respect Dr. Maynard did not fully realize the effect of his own hypothesis. After suggesting that the effect of inoculation was simply to shift the distribution of resistance to the right along the scale of "strength to resist attack," he merely concludes that "considering these difficulties . . . we cannot consider the 'fourfold' method as entirely satisfactory for our present purpose; nevertheless, if a correlation value is desired it is, I believe, the best method at present available for this purpose," and later remarks of the method that "although not perfect for this purpose it is useful as a control." Such observations read as if Dr. Maynard merely supposed that, on his hypothesis, the normal coefficient became a rather poor approximation to the true value of the correlation. But the coefficient appears to us to have become quite meaningless, for it purports to give a correlation which *cannot* be determined from the given data.

(*B*) *The Product-sum Correlation.*—This is the correlation coefficient (cf. Yule, 1912, pp. 595 *et seq.*)—

$$r = \sqrt{\chi^2/N}$$
$$= \sqrt{(p_1 - p_2)\ (p_3 - p_4)}$$

where p_3 and p_4 are the proportions in the columns corresponding to p_1 and p_2 in the rows. Of this coefficient it is enough to remark that it varies with the number of the vaccinated even more rapidly than the normal coefficient. For the values of the fatality-rates assumed in the illustration given for that coefficient, the product-sum correlation falls from 0·35 when the proportion of vaccinated is 50 per cent. to 0·06 when the proportion is 1 per cent.

(*C*) *The Coefficient of Colligation* (Yule, 1912).—This coefficient, suggested by one of us a few years ago, may be expressed in terms

of the class-frequencies a, b, c, d, or in terms of p_1 and p_2, or in terms of p_3 and p_4 as follows:—

$$\omega = \frac{\sqrt{ac} - \sqrt{bd}}{\sqrt{ac} + \sqrt{bd}}$$

$$= \frac{\sqrt{p_1(1-p_2)} - \sqrt{p_2(1-p_1)}}{\sqrt{p_1(1-p_2)} + \sqrt{p_2(1-p_1)}}$$

$$= \frac{\sqrt{p_3(1-p_4)} - \sqrt{p_4(1-p_3)}}{\sqrt{p_3(1-p_4)} + \sqrt{p_4(1-p_3)}}$$

Since the coefficient can be expressed as a function of p_1 and p_2 only, it is evidently quite independent of the proportion of the population that is vaccinated, and is, therefore, not open to objection on the same grounds as the normal coefficient and the product-sum correlation. As a matter of fact, if the actual frequencies in the rows and columns of the observed table are multiplied by such coefficients that not only is the number of vaccinated made equal to the number of unvaccinated but also the number of recoveries made equal to the number of deaths, then the coefficient ω is the product-sum correlation for this derived symmetrical table. If the square-roots are omitted in the expression for ω, an even simpler expression (Q) is obtained which was given as a coefficient of association by one of us in 1900; it enjoys the same property as ω in being independent of alterations in the row or column totals, but is not so simply related to the product-sum coefficient and gives a less convenient scale.

Is this coefficient, then, any better than the others, seeing that at least it passes the elementary test? The answer must be, it is no better. It is quite right that the coefficient should be unaffected by altering the proportion of the vaccinated. For if we double or treble the area of tne curve representing the distribution for the vaccinated in fig. 8 this will not (of course) alter the proportions of dead cut off to the left of the verticals 1, 2, 3, &c., in each curve. But the coefficient is *also* unaltered by varying the proportions of recoveries and deaths in the population, for this will leave the p_3 and p_4 of the table—the proportions of vaccinated amongst recoveries and deaths—unaltered, and this is *not* right on the theory developed in the last section. For if we alter the proportion of recoveries by shifting the critical vertical to the right, we *do* alter the proportions of vaccinated on the right of that vertical (amongst the recoveries) and on the left of that vertical (amongst the deaths), and we cannot help doing so. The coefficient is not constant therefore for varying inten-

sities of epidemic. The following table, calculated from the curves of fig. 8, shows the values of ω that would be obtained on our theory

Proportion of recoveries amongst uninoculated	Coefficient of colligation
0·1	0·7071
0·2	0·6625
0·3	0·6327
0·4	0·6188
0·5	0·6050
0·6	0·5934
0·7	0·5817
0·8	0·5691
0·9	0·5477

for successive proportions of recoveries amongst the uninoculated in the case of the cholera data utilized in illustration A of the last section. It will be seen that the coefficient falls steadily as the virulence of the epidemic decreases. There is no reason, of course, why a measure of efficiency should be constant over the range—we have seen in fact above that such is not the case; but the point is that we can assign no interpretation to the change, seeing that the coefficient assumes as a possibility what is, on our view of the facts, impossible. For the same disease and the same type of inoculation it is not, on our theory, possible to alter the proportion of deaths without altering the proportions of vaccinated among deaths and recoveries. The same arguments suffice to condemn the similar coefficient, Q.

This brief review does not, of course, cover all the possible coefficients that have been suggested, but will suffice to confirm the argument of this section. Our condemnation of these coefficients, for the present purpose, may at least claim to be impartial, inasmuch as two of them were originally proposed by one of us.

The results of this section are disappointing in so far as they fail to provide a simple answer to important practical questions. On the other hand, we venture to hope that they will be of value to subsequent inquirers. The general lesson to be learned is that mathematical difficulties of method must not absorb the whole energies of the statistician. To Professor Pearson and his pupils we owe the solution of many mathematical difficulties, but Dr. Brownlee and Dr. Maynard alone, so far as we are aware, have assigned a due measure of importance to the biological difficulties of interpretation which present themselves in connexion with such inquiries as that on which we have been engaged.

BIBLIOGRAPHY.

Antityphoid Committee, Report of, 1912.

BROWNLEE, J. "Statistical Studies in Immunity," *Biometrika*, 1905, iv, p. 313.

Idem. "The Causes of Epidemics and the Laws which regulate their Course," *Proc. Sect. XVIII Internat. Congress of Med.*, 1913.

Idem. "On the Curve of an Epidemic," *Brit. Med. Journ.*, 1915, i, p. 799.

CARDAMATIS, J. P. "Rapport sur la lutte contre le choléra en Macédoine pendant la guerre Gréco-Bulgare," *Bull. de la Soc. de Path. Exot.* 1914, vii, p. 447.

FIRTH, R. H. "A Statistical Study of Anti-enteric Inoculation," *Journ. Roy. Army Med. Corps*, 1911, xvi, p. 589.

GREENWOOD, M., jun. "The Factors that determine the Rise, Spread, and Degree of Severity of Epidemic Disease," *Proc. Sect. XVIII Internat. Congress of Med.*, 1913.

HAFFKINE, W. M. "Protective Inoculation against Cholera," Calcutta, 1913.

MACDONNELL, W. R. "On the Influence of Previous Vaccination in Cases of Small-pox," *Biometrika*, 1901, i, p. 375; 1902, ii, p. 135.

MAYNARD, G. D. "Statistical Studies of Anti-typhoid Inoculation," *Biometrika*, 1908, vi, p. 366.

Idem. "An Enquiry into the Etiology, Manifestations, and Prevention of Pneumonia amongst Natives on the Rand recruited from Tropical Areas," Johannesburg, 1913. (South African Institute for Medical Research.)

MOUTOUSES, K. J. "Die Cholerabekämpfung i. d. griechischen Armee während des griechischen-bulgarischen Kriegs," *Militärarzt*, 48 J., p. 66. (Supplement to No. 10, *Wien. med. Wochenschr.*, 1914.)

MURATA, N. "Ueber die Schutzimpfung gegen Cholera," *Centralbl. f. Bakt.* (Orig.), 1904, xxxv, p. 605.

NIJLAND, A. H. "Eenige resultaten met het Cholera-vaccin verkregen," *Geneeskundig Tijdschrift v. Nederlandsch-Indië*, 1911, Deel 51, afl. 4, and 1913, Deel 53, afl. 1.

PEARSON, K. "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling," *Phil. Mag.*, July, 1900.

Idem. "On the Correlation of Characters not Quantitatively Measurable," *Phil. Trans.*, A, 1900, cxcv, p. 1.

PEARSON, K., and HERON, D. "On Theories of Association," *Biometrika*, 1913, ix, p. 159.

POWELL, A. "Results of Haffkine's Anti-cholera Inoculation," *Lancet*, 1896, ii, p. 171.

Idem. "Further Results of Haffkine's Anti-cholera Inoculation," *Journ. Trop. Med.*, 1899, vi, p. 115.

ROSS, Sir R. "Some *a priori* Pathometric Equations," *Brit. Med. Journ.*, March 27, 1915, p. 546.

SAVAS, C. "Ueber die Choleraschutzimpfung in Griechenland," *Wien. klin. Wochenschr.*, 1914, xxvii, p. 1093.

SIMON, P. L.; PASTEUR VALLERY-RADOT; KIAMIL BEY; ASSEO. "Notes sur le Choléra à Constantinople et en Thrace de 1910 à 1913," *Bull. de la Soc. de Path. Exot.*, 1914, vii, p. 313.

STICKER, C. "Die Cholera," Giessen, 1912.

Idem. "Ueber die Bedingungen, von denen die Entwicklung, die Ausbreitung und die Schwere der Epidemien abhängen," *Proc. Sect. XVIII Internat. Congress of Med.*, 1913.

YULE, G. U. "On the Association of Attributes in Statistics," *Phil. Trans.*, A, 1900, cxciv, p. 257.

Idem. "On the Methods of Measuring Association between two Attributes," *Journ. Roy. Stat. Soc.*, 1912, lxxv, p. 579.

ZABOLOTNY. "Schutzimpfung gegen Cholera," *Bericht u. d. 14ten Congress f. Hygiene z. Berlin*, 1907.

Idem. "L'Epidémie du Choléra 1907-08 en Russie," *Arch. d. Sci. Biol.*, St. Petersb., 1909, xiv.

DISCUSSION.

Dr. W. H. HAMER said he was very sorry to be occupying the Chair, as his doing so was caused by the absence of the President; but he was very glad to have enjoyed the pleasure of being present, when the extremely interesting paper they had just listened to was read. The great advances which were being made year by year in the sciences of bacteriology and statistics made the lot of the medical officer of health anything but a happy one. On the last occasion when he was present at a meeting of the Section, they were given marching orders with regard to cerebrospinal meningitis. It was a comparatively simple matter to find a meningococcus carrier, but what was to be done with him when he was found. However, he would not enlarge upon that topic, for fortunately the outbreak of cerebrospinal fever was abating, and clearly he had better not attempt to estimate how far this was due to restraining the movements of healthy carriers, and how far it was due to the natural course of events. With regard to typhoid prevention, two points connected with typhoid fever in London particularly struck him: *first*, the immunity of the London population from the disease—in one or two recent weeks the number of cases had fallen as low as five, a very small number for so large a community; and, *secondly*, even now, there was very great uncertainty as to the diagnosis of typhoid fever, for the percentage of errors of diagnosis in cases sent to the Metropolitan Asylums Board Hospitals amounted to nearly forty. If such difficulties with regard to diagnosis were experienced in London, he wondered whether similar difficulties might not also be experienced elsewhere. One question he would particularly like to ask Mr. Greenwood, as it possibly might have a bearing upon the statistical theory developed in Section III of the paper. In the case of swine fever, inoculation was now being practised with blood which it was presumed contained the filter-passer cause of the disease: some work had been done, however, in years gone by, with cultures of the hog cholera bacillus, an organism which was at that earlier period actually believed to be the cause of hog cholera. It would be extremely interesting, if it were possible, to make comparison between the results obtained by employing these two methods: *first*, in respect of protection against attack; and *secondly*, in respect of protection against a fatal result. It might, perhaps, be suggested, on merely theoretical grounds, that cultures of an "associated organism" were more likely to protect against a fatal result than against attack by the disease; while inoculation with the filter-passing causal organism might perhaps be expected to protect, if at all, rather against attack than against severity of illness. If Mr. Greenwood, who was particularly interested in the subject of swine fever, could give them any enlightenment on these points, the speaker was sure they would all be grateful, for if it was possible to investigate the facts, so far as swine fever was concerned, on these lines, the results obtained might then be considered in their bearing upon the very difficult case of typhoid fever. He

was assuming that everything was not at the present time known about the causal organism of typhoid fever, a supposition which some high authorities seemed inclined to doubt.

The discussion was continued by Dr. E. W. GOODALL, Dr. BROWNLEE, Dr. G. S. BUCHANAN, and Dr. J. C. MCVAIL.

Professor W. J. R. SIMPSON, C.M.G., said that the authors were to be congratulated on their very able paper. The array of assumptions and figures in the paper were perplexing to one who was not a mathematician, but it was obvious to all that Mr. Major Greenwood and Mr. Yule had approached the subject with an open and thoroughly critical mind and had dealt with it in a most comprehensive manner. He was glad to find that the result of their examination of the statistics was the conclusion that the Calcutta, Assam, and Java statistics, together with those relating to the Sanitary Corps of the Greek Army, afforded sufficient ground for the introduction of cholera inoculations in the British Army as a prophylactic against cholera. Though cautiously expressed, it was an important pronouncement, and one with which he agreed. In fact, his own experience led him to advocate the prophylactic measure much more strongly than the writers of the paper. He thought that no time should be lost in protecting the soldiers, especially in Egypt and the Dardanelles, against this disease. He would like to point out that owing to the favourable conditions under which the Calcutta observations were made and the special care taken to secure their accuracy, the figures, though small in numbers, were more significant than they appeared, for they represented over seventy distinct and separate experiments, not during the period of a single epidemic but extending over more than two years in an endemic locality. Cholera was endemic in Calcutta. It was more prevalent in the early and later months of the year than at other times, but there was no month in which the town was free of the disease. There were certain parts of the town in which the disease recurred more frequently than in others, and it was here that most of the 7,000 inoculations were carried out. The method adopted was as far as possible to inoculate half the members of a household and leave the other half not inoculated, and then await events. During the period under observation cholera occurred in many of the houses where no inoculations had been performed. These cases were of no special interest. But cholera also entered some seventy houses where part of the household had been inoculated and the remainder had not been inoculated. It was remarkable to note how, after the fourth day of the inoculation, when immunity had been established among the inoculated, cholera picked out in these houses the non-inoculated and left the inoculated unharmed. This went on repeating itself month after month to such an extent that everyone connected with watching the experiments was deeply impressed by the protective value of the inoculations, the protection lasting for over fourteen months. Even the inhabitants of these areas noticed the difference of the incidence of cholera among the inoculated and non-inoculated and came forward to be inoculated, many insisting on all the members of the household being given the prophylactic.

Mr. YULE, replying to some of the theoretical points raised in the discussion, said that he had listened with much interest to the remarks made by Dr. Brownlee and was glad to know that that speaker was in general agreement with their views. He well remembered Dr. Brownlee's statement as to the distinctness of small-pox in the case of vaccinated and unvaccinated patients. On this point Mr. Greenwood and himself were rather inclined to agree with Dr. Buchanan: if the diseases in the two cases were almost specifically distinct their hypothesis would hardly apply. They must be sufficiently similar to enable one to plot the two curves of resistance to the same scale—i.e., the difference must be of a quantitative rather than of a qualitative kind. He would like to know more of Dr. Brownlee's views on this head. Dr. Brownlee had also called attention to the frequency of linear regression in statistical work. They agreed that such regression was frequent, and might perhaps add that, for this reason, they could not regard the closeness with which their hypothesis was fulfilled as any strong evidence of its truth. The hypothesis must at present rest rather on the *a priori* evidence that it was a natural and reasonable assumption than on any support afforded by statistics. Professor Simpson had referred to the caution of their conclusions as regards cholera inoculation, and the conviction that had been carried to his own mind by actual experience when he had been able to note the experience of inoculated and uninoculated persons in the same house. They had been compelled, of course, to deal with the data in a summary way and to point out the imperfections of many of the data presented, but he did not think that there was any doubt about their final conclusion, especially when the data of Savas and of Haffkine were considered together. Where they did find themselves in real difficulties was on a different point—namely, when they tried to compare the efficiency of the immunization process in the case of cholera with that of the immunization process in the case of another disease. Of this they were unable to provide a statistical measure. With the remarks made by Dr. McVail he was in complete agreement. Dr. Buchanan had commented on a passage respecting the distribution of "effective vaccination." It was, of course, not quite clear what Professor Pearson understood by that term: it had to be interpreted in the sense of a variable magnitude that could be plotted on a scale, and it then seemed to them to be almost identical with resistance. In any case, it did seem to them that a person with no cicatrix recorded might be more "effectively vaccinated" than a person with a very faint cicatrix; or, if he had had a previous attack of small-pox, than a person even with a well-marked cicatrix. Mathematicians were just as likely to be wrong as other people were, and data drawn from any particular science could not be adequately treated by the mathematician unless he was as much of an expert in that particular science as he was in mathematics. He himself would not have attempted to write such a paper as the present alone; fortunately, he had been associated with Mr. Greenwood.

Mr. MAJOR GREENWOOD, jun., desired merely to express his complete concurrence with respect to points his colleague had mentioned in replying

on the discussion. He might emphasize Mr. Yule's remark that they both intended to convey the conclusion that, so far as the data went, the case in favour of anti-cholera inoculation seemed to them strong. The question as to which arrangement of Haffkine's material should be chosen had been considered, and he believed that that adopted in the paper was the most satisfactory. A separate consideration of the incidence of disease upon individual houses would not, he thought, have led to the expression of any more pointed conclusions.