PRINCIPLES OF MEDICAL STATISTICS

XII—COMMON FALLACIES AND DIFFICULTIES

IN the following sections I have set out examples of the misuse of statistics. In some of them the actual figures have been taken from published papers; in others hypothetical figures have been used to indicate the type of error which has led the worker to fallacious conclusions. No principles are involved that have not been discussed in the previous sections. The object is merely to illustrate, at the risk of "damnable reiteration," the importance of these principles by means of simple numerical examples (in some instances—e.g., (*a*) below—the figures are deliberately exaggerated to make clearer the point at issue). The fact that in practice such grossly exaggerated differences rarely occur does not lessen the importance of accurate statistical treatment of data. Differences do occur very often of a magnitude to lead to erroneous conclusions, if the data are incompetently handled in the ways set out.

Mixing of Non-comparable Records

(*a*) Let us suppose that in a particular disease the fatality-rate is twice as high amongst females as it is amongst males, and that amongst male patients it is 20 per cent. and amongst female patients 40 per cent. A new form of treatment is adopted and applied to 80 males and 40 females; 30 males and 60 females are observed as controls. The number of deaths observed amongst the 120 *individuals* given the new treatment is 32, giving a fatality-rate of 26·7 per cent., while the number of deaths observed amongst the 90 *individuals* taken as controls is 30, giving a fatality-rate of 33·3 per cent. Superficially this comparison suggests that the new treatment is of some value; in fact that conclusion is wholly unjustified, for we are not comparing like with like. The fatality-rates of the total number of individuals must be influenced by the proportions of the two sexes present in each sample; males and females, in fact, are not equally represented in the sample treated and in the sample taken as control. Tabulating the figures shows the fallacy clearly (Table XIV).

TABLE XIV

—	Males.	Females.	Males and females combined.
Normal fatality-rate ..	20 %	40 %	—
Number of patients given new treatment	80	40	120
Deaths observed in treated group	16	16	32
Fatality-rates observed in treated group	20 %	40 %	26·7 %
Number of patients used as controls	30	60	90
Deaths observed in control group	6	24	30
Fatality-rates observed in control group	20 %	40 %	33 %

The comparison of like with like—i.e., males with males and females with females—shows that the treatment was of no value since the fatality-rates of the treated and untreated sex groups are identical, and equal to the normal rates. Comparison of the total samples, regardless of sex, is inadmissible for the fatality-rate recorded is then in part dependent upon the proportion of the two sexes that are present. There are proportionately more females amongst the controls than in the treated group, and since females normally have a higher fatality-rate than males their presence in the control group in relatively greater numbers must lead to a comparatively high fatality-rate in the total sample. Equally their relative deficiency in the treated group leads to a comparatively low fatality-rate in that total sample. No comparison is valid which does not allow for the sex differentiation of the fatality-rates.

An actual example of this error is in the record of a less favourable reaction of females to forms of treatment for syphilis. It should not be forgotten that when two such forms of treatment are being compared, sex differentiation, in assessing the results (or sex equality in the two groups), may be of importance.

(*b*) One more example of the result of mixing non-comparable records may be given. The following (hypothetical) figures show the attack-rates of a disease upon an inoculated and an uninoculated population (Table XV).

TABLE XV

Year.	Number of persons.		Number of persons attacked.		Attack-rates per cent.	
	Inoc.	Uninoc.	Inoc.	Uninoc.	Inoc.	Uninoc.
1935 ..	100	1000	10	100	10	10
1936 ..	500	600	5	6	1	1
1935 and 1936	600	1600	15	106	2·5	6·6

In each calendar year the attack-rate of the inoculated is equal to the attack-rate of the uninoculated. Between 1935 and 1936 there has, however, been a large change in the size of the inoculated and uninoculated population and also a large change in the level of the attack-rate. Summation of the results for the two years leads to the fallacious conclusion that the inoculation afforded some protection. The large uninoculated population in 1935 when the attack-rate was high leads to an absolutely large number of cases—though in relation to their numbers the uninoculated are at no disadvantage compared with the inoculated. The inoculated cannot contribute an equal number of cases for the population at risk in that year (1935) is far smaller. Thus amalgamation of the unequal numbers of persons exposed to

TABLE XVI

Year.	Number of persons.		Number of persons attacked.		Attack-rates per cent.	
	Inoc.	Uninoc.	Inoc.	Uninoc.	Inoc.	Uninoc.
1935 ..	100	1000	10	100	10	10
1936 ..	500	600	50	60	10	10
1935 and 1936	600	1600	60	160	10	10
1935 ..	500	600	50	60	10	10
1936 ..	500	600	5	6	1	1
1935 and 1936	1000	1200	55	66	5·5	5·5

different risks in the two years is unjustified. No fallacy would have resulted if the attack-rate had not changed or if the proportions exposed to risk had not changed, as the above figures show (Table XVI).

When the populations at risk and the attack-rates *both* vary, the calendar year becomes a relevant factor, and must be taken into account by the calculation of rates within the year.

Such a problem does arise quite frequently in practice, for example in assessing the incidence of diphtheria on immunised and unimmunised children. The numbers in the immunised group vary from year to year and the incidence of diphtheria also varies. Summation of the experience over a series of years may lead to an erroneous conclusion of the type illustrated.

Neglect of the Period of Exposure to Risk

(*a*) A further fallacy in the comparison of the experiences of inoculated and uninoculated persons lies in neglect of the time during which the individuals are exposed first in one group and then in the other. Suppose that in the area considered there were on Jan. 1st, 1936, 300 inoculated persons and 1000 uninoculated persons. The number of attacks are observed within these two groups over the calendar year and the annual attack-rates are compared. This is a valid comparison *so long as the two groups were subject during the calendar year to no additions or withdrawals.* But if, as often occurs in practice, persons are being inoculated *during* the year of observation the comparison becomes invalid unless the point of time at which they enter the inoculated group is taken into account.

Suppose on Jan. 1st, 1936, there are 5000 persons under observation, none of whom are inoculated; that 300 are inoculated on April 1st, a further 600 on July 1st, and another 100 on Oct. 1st. At the end of the year there are, therefore, 1000 inoculated persons and 4000 still uninoculated. During the year there were registered 110 attacks amongst the inoculated persons and 890 amongst the uninoculated. If the ratio of recorded attacks to the population *at the end of the year* is taken, then we have rates of 110/1000=11·0 per cent. amongst the inoculated and 890/4000=22·3 per cent. amongst the uninoculated, a result apparently very favourable to inoculation. This result, however, *must* be reached even if inoculation is completely valueless, for no account has been taken of the unequal lengths of time over which the two groups were exposed. None of the 1000 persons in the inoculated group were exposed to risk for the *whole* of the year but only for some fraction of it; for a proportion of the year they belong to the uninoculated group and must be counted in that group for an appropriate length of time.

The calculation should be as follows:

All 5000 persons were uninoculated during the first quarter of the year and therefore contribute $(5000\times\frac{1}{4})$ years of exposure to that group. During the second quarter 4700 persons belonged to this group—i.e., 5000 less the 300 who were inoculated on April 1st—and they contribute $(4700\times\frac{1}{4})$ years of exposure to the uninoculated group. During the third quarter 4100 persons belonged to this group—i.e., 4700 less the 600 who were inoculated on July 1st—and they contribute $(4100\times\frac{1}{4})$ years of exposure. Finally in the last quarter of the year there were 4000 uninoculated persons—i.e., 4100 less the 100 inoculated on Oct. 1st—and they contribute $(4000\times\frac{1}{4})$ years of exposure. The "person-years" of exposure in the uninoculated group were therefore $(5000\times\frac{1}{4})+(4700\times\frac{1}{4})+(4100\times\frac{1}{4})+(4000\times\frac{1}{4})=4450$, and the attack-rate was 890/4450=20 per cent.—i.e., the equivalent of 20 attacks per 100 persons per annum. Similarly the person-years of exposure in the inoculated group are $(0\times\frac{1}{4})+(300\times\frac{1}{4})+(900\times\frac{1}{4})+(1000\times\frac{1}{4})=550$, for there were no persons in this group during the first three months of the year, 300 persons during the second quarter of the year, 900 during the third quarter, and 1000 during the last quarter. The attack-rate was, therefore, 110/550 = 20 per cent., and the inoculated and uninoculated have identical attack-rates. Neglect of the durations of exposure to risk must lead to fallacious results and must favour the inoculated. The figures are given in tabulated form (Table XVII).

TABLE XVII

Inoculated at each point of time.	Inoculated.		Uninoculated.	
	Exposed to risk in each quarter of the year.	Attacks at 5 per cent. per quarter.	Exposed to risk in each quarter of the year.	Attacks at 5 per cent. per quarter.
Jan. 1st, 0	0	0	5000	250
Apr. 1st, 300	300	15	4700	235
July 1st, 600	900	45	4100	205
Oct. 1st, 100	1000	50	4000	200
Total at end of the year ..	1000	110	4000	890

Fallacious comparison.—Ratio of attacks to final population of group. Inoculated 110/1000=11·0 per cent. Uninoculated 890/4000=22·3 per cent.

True comparison.—Ratio of attacks to person-years of exposure. Inoculated $110/(300\times\frac{1}{4})+(900\times\frac{1}{4})+(1000\times\frac{1}{4})=20$ per cent. Uninoculated $890/(5000\times\frac{1}{4})+(4700\times\frac{1}{4})+(4100\times\frac{1}{4})+(4000\times\frac{1}{4})=20$ per cent.

This example is an exaggerated form of what may (and does) happen in practice if the time-factor is ignored. Clearly even if the time-factor is allowed for, interpretation of the results must be made with care. If the inoculated show an advantage over the uninoculated it must be considered whether at the point of time they entered that group the incidence of the disease was already declining, due merely to the epidemic swing. But that is another point.

(*b*) A cruder neglect of the time-factor sometimes appears in print, and may be illustrated as follows. In 1930 a new form of treatment is introduced and applied to patients seen between 1930 and 1935. The proportion of patients still alive at the end of 1935 is calculated. This figure is compared with the proportion of patients still alive at the end of 1935 who were treated in 1925–29, prior to the introduction of the new treatment. Such a comparison is, of course, inadmissible. The patients seen in 1925–29 have by the end of 1935 had 6 to 11 years in which to succumb, with an average exposure of 8½ years if their attendances were equally spread over 1925–29. The patients seen in 1930–35 have had only 0 to 6 years in which to succumb, with an average exposure of 3 years if there attendances were equally spread over 1930–35. To be valid the comparison must be between the survival-rates at equal stages of time, 1, 2, 3 years, &c., after treatment.

Absence of Exposed to Risk or Standard of Comparison

It often happens that an investigation is confined to individuals marked by some characteristic.

(*a*) For example a detailed inquiry is made into the home conditions of each infant dying in the first year of life in a certain area over a selected period of time, and it is found that 15 per cent. of these infants lived under unsatisfactory housing conditions. Do such conditions, or factors associated with them, lead to a high rate of infant mortality? The limitation of the inquiry to the dead makes it quite impossible to answer this

question. We need information as to the proportion of *all* infants who were born in that area over that period of observation who live under unsatisfactory housing conditions. If 15 per cent. of *all* infants live under such conditions, then 15 per cent. of the deaths may reasonably be expected from those houses and unsatisfactory housing appears unimportant. If on the other hand only 5 per cent. of *all* infants are found in these conditions but 15 per cent. of the deaths come from such houses, there is evidence of an excess of mortality under the adverse conditions. In practice it may be impossible for financial or administrative reasons to investigate the home conditions of all infants. It should be possible, however, to inquire into a random sample of them, say every tenth birth registered in the area over a given period of time. Without some such standard of comparison no clear answer can be reached. Such limited investigations have been made into the problems of both infant and maternal mortality.

(*b*) After very careful inquiry it is shown that of motor-car drivers involved in accidents a certain proportion, say three-quarters, had consumed alcohol during some period of hours previous to the accidents, and one-quarter had not. The deduction that alcohol contributes to the risk of accident is not justified from these figures alone. It is well recognised that white sheep eat more than black sheep —because there are more of them. Before the ratio of 3 "alcoholics" to 1 "non-alcoholic" amongst the accident cases can be interpreted, information is also required as to the comparable ratio amongst drivers *not involved in accidents*. Suppose, for example, there are 1000 drivers on the roads, and 48 accidents are recorded. Of the 48 drivers involved in these accidents three-quarters are found to have consumed alcohol—i.e., 36—and one-quarter—i.e., 12—have not. If three-quarters of *all* the 1000 drivers have consumed alcohol within a few hours of driving and one-quarter have not, then the populations "exposed to risk" of accident are 750 and 250. The accident-rates are, then, identical—namely, 36 in 750 and 12 in 250, or 4·8 per cent. in each group. A knowledge of the exposed to risk, or of the ratio of alcohol consumers to non-consumers in a random sample of all drivers, is essential before conclusions can be drawn from the ratio in the accident cases.

Careful inquiry into the destination of drivers involved in accidents on a Sunday morning might show that a larger proportion was driving to golf than to church. The inference that driving to golf is a more hazardous occupation is not valid until we are satisfied that there are not, in this case, more black sheep than white sheep.[1]

Association, Direct or Indirect

It has been observed that while the death-rate from cancer has been rising the sale of bananas in England and Wales has also been increasing. No one (so far as I know) has deduced from this relationship in time a relationship of cause and effect. But such a deduction would be no less logical than many that are drawn from time relationships—e.g., the relationship of the cancer death-rate to a hundred-and-one aspects of "modern" life.

More often the time relationship is used as further evidence of a cause and effect expected on the grounds of quite other evidence—e.g., experimental results.

(*a*) For instance, the effect of bacteriophage is measured by comparing the incidence of cholera in two areas, one in which bacteriophage was distributed, the other serving as a standard of comparison. In the former the incidence is found over an observed period of time to be at a lower level than the incidence in previous years or in the area observed as a control (the question of duration of exposure to risk, dealt with above, having been properly observed). It is clear that there is an association both in time and space between the incidence of cholera and the administration of bacteriophage. Is that association one of cause and effect? The answer must be that the results are perfectly *consistent* with that hypothesis, but that consistency is not the equivalent of proof. The incidence of epidemic disease fluctuates both in time and space for unknown reasons, and the abnormally low attack-rates in the area in which bacteriophage was administered *may* be the result of the influence of those undetermined natural causes operating at the same time as the experiment was carried out. Repetition of the experiment in another area with equivalent results would strengthen the hypothesis that bacteriophage was beneficial. With observations of this kind, limited in time and space, it is well to reflect upon the fact that "if when the tide is falling you take out water with a twopenny pail, you and the moon together can do a great deal." The history of scarlet fever may well be remembered, in this connexion, as illustrated by the testimony of R. J. Graves ("A System of Clinical Medicine," Dublin, 1843). In the first few years of the nineteenth century the disease "committed great ravages in Dublin" and was "extremely fatal." After the year 1804 it assumed a "very benign type" and was "seldom attended with danger until the year 1831." In 1834 it again took the form of a "destructive epidemic." The low fatality after 1804 was "every day quoted as exhibiting one of the most triumphant examples of the efficacy" of new methods of treatment. But Graves candidly admits that "the experience derived from the present [1834–35] epidemic has completely refuted this reasoning, and has proved that, in spite of our boasted improvements, we have not been more successful in 1834–5 than were our predecessors in 1801–2" (quoted from Charles Creighton's "History of Epidemics in Britain," Camb. Univ. Press, 1894, Vol. II, pp. 722–25).

(*b*) During an epidemic individuals are, *at their own request*, inoculated, and are found to suffer a lower attack-rate than that of the uninoculated. This result is clearly consistent with the hypothesis that the inoculation was beneficial. On the other hand, the association may be an indirect one. Presuming that the inoculated and uninoculated groups were equal in such characteristics as age, sex, and duration of exposure to risk, it is possible that they were differentiated in other relevant ways. Those who voluntarily come forward for inoculation may consist of individuals who also take other precautionary measures to avoid infection—e.g., the avoidance of theatres and cinemas during the epidemic period. They may belong to a class higher in the social scale and be less exposed to the risk of infection thereby—e.g., by living in less crowded conditions or by being better nourished. In comparing the inoculated with the uninoculated we must always consider closely whether we are in fact comparing like with like—except in the one respect of inoculation.

A. B. H

[1] Lest it should be thought that undue stress is being laid upon the obvious, the following quotation from a debate in the House of Lords may be of interest. A noble Lord is reported (*Times*, Feb. 7th, 1936) to have said that "only 4 per cent. of the drivers involved in fatal accidents were women, and that was because they drove more slowly." Without evidence of the hours of driving endured (perhaps a fitting word nowadays) by each sex—and perhaps of the type of area—that conclusion cannot be justified.