Some Aspects of Retrospective Studies

15 DE NOVEMBRO DE 1975 — SÁBADO — 8:00 h

Dr. William Haenszel

THE statement that a disease and a characteristic are associated can be interpreted in a number of different ways.

(1) The deliberate introduction of the characteristic into a subgroup of a population will be followed by an altered incidence of the disease in that subgroup.

(2) Those members of a population possessing the characteristic will experience a different incidence of the disease than will those lacking it, the route by which the characteristic was acquired being unspecified.

(3) The prevalence of the characteristic among newly developed cases of the disease differs from its prevalence among those in whom the disease did not develop.

The logical content of these three statements is obviously not the same, and the presence of an association of any one of the three types does not necessarily imply that a similar association exists for the other two. We may refer to the first as a causal association and to the other two as observational associations.

In studies of disease etiology, the major interest is in causal associations. It is often not possible, however, to study such associations by direct experimentation in populations of interest, especially when the introduced characteristic may lead to an increased incidence of the disease. The introduction of control measures designed to decrease the incidence of the disease will, of course, sometimes provide such an experimental test.[1] When experimental tests are not possible, however, inferences about causal associations must be drawn either from a variety of observational associations or from a study of causal associations in animal populations or from a combination of both. In the nature of the case such inferences cannot be certain and, what is perhaps worse, no quantitative characterization of the uncertainty appears possible. There appear to be no general rules of inference or routines of thought that are helpful in such situations, although a number have been proposed. Particularly mischievous in this connection, in our opinion, is the distinction sometimes drawn between "statistical" and "biologic" evidence.

MEASUREMENT OF RISK OF DISEASE OCCURRENCE

Our concern here, however, is not with these questions of inference but with observational associations, the study of which is clearly required, irrespective

of the exact nature of the causal inferences that they will support. Lacking the ability to introduce deliberately the characteristic into a subgroup, the next best step involves subclassifying the members of some well-defined population

TABLE I

CHARACTERISTIC	DEVELOPMENT OF DISEASE		TOTAL
	NUMBER OF INDIVIDUALS +	NUMBER OF INDIVIDUALS −	
+	a	b	a + b
−	c	d	c + d
Total	a + c	b + d	N

(or a sample of it) according to whether they do or do not possess the characteristic. One may then count the number of cases of disease for those with and without the characteristic in a number of different ways: (1) the number of new cases of the disease developing during some period of time, say a year, usually designated by the phrase *incidence*, (2) the number of deaths from the disease occurring during a given period of time, the *mortality*, (3) the number of live individuals having the disease at some moment of time, the *point prevalence*, or (4) the number of live individuals having the disease at any time during some interval, the *interval prevalence*. Certain systematic relationships obtain among these different measures.[2] The interval prevalence of a disease, for example, will never be less than any point prevalence within that interval. In comparisons among groups with different characteristics, the various measures will often yield quite similar results. Although strictly speaking etiological aspects are best analyzed from the point of view of incidence, these empirical relationships have encouraged the use of the most convenient measure for study purposes, usually mortality. In strict logic, however, there is no reason why the different measures should yield even qualitatively similar results. Neyman, for instance, has constructed an example in which comparison of point prevalence in two groups showed a large difference in one direction, while comparison of incidence showed a large difference in the opposite direction.[3] Furthermore, examples are sometimes encountered in practice in which this question assumes importance. Thus, because tuberculosis has a more rapidly fatal course among Negroes than among white persons, the prevalence rates in the two groups are approximately equal, even though the incidence of the disease is greater in Negroes than in white persons.[4]

In the discussion that follows, we shall assume that the measure adopted is incidence. With certain obvious modifications in language, everything said

will be applicable to other measures as well; the wisdom of using another measure in any particular instance depends upon special circumstances which must be individually evaluated.

A population may be classified by presence or absence of the characteristic and development or nondevelopment of the disease (Table I). The number of individuals among those with the characteristic newly developing the disease is designated by a; b designates the number with the characteristic that did not develop the disease; and so forth. The incidence of the disease for the time period covered among those with the characteristic is thus $\frac{a}{a+b}$ and among those without it $\frac{c}{c+d}$. It is customary to speak of the relative risk of developing the disease, defined as the ratio of the incidence for those with and without the characteristic:

$$\text{Relative risk} = \frac{a}{a+b} \Big/ \frac{c}{c+d}$$

Thus, if $a = 1{,}000$, $b = 99{,}000$, $c = 500$, and $d = 99{,}500$, the risk of developing the disease for those with the characteristic is twice the risk of those lacking it. If this relative risk is unity, we say the disease and characteristic are unassociated. The value of the relative risk, when other than unity, provides a measure of the degree of association; ratios in excess of unity indicate positive, and below unity negative, association. Other measures of association are possible, of course, but in the early stages of an investigation of the causal role of a characteristic, the relative risk provides a most useful descriptive summary of the association. A more extended discussion of this viewpoint will be presented in the section on "Potential Sources of Error in Retrospective Studies." Other discussions of interest are given by Berkson,[5] Sheps,[6] and Goodman and Kruskal.[7]

RELATION BETWEEN PROSPECTIVE AND RETROSPECTIVE STUDIES

Studies which start with populations grouped initially into subclasses, for each of which one counts the number of new cases of a disease which develop during some subsequent period of time, are ordinarily referred to as "prospective" or "population-based" studies. The annual incidence of most diseases is sufficiently small, so that prospective studies designed to supply estimates of the incidence rate for different classes of the population, or of their ratios, must cover large numbers of persons. Thus, in a prospective study of lung cancer in a population of 100,000 males over age 40, one might at the end of 1 year of study expect to find 50 to 75 new cases. This is a small return for a large effort. The "retrospective" or "case-control" study provides a more economical way of estimating the relative risk than the prospective method because it does not require devotion of a large part of the study resources to those who did not develop the disease. In such a study one identifies all, or a well-defined sample, of the new cases of a disease as they occur during some period of time, and only after the occurrence

of the disease does one classify them by the presence or absence of the characteristic (hence the name "retrospective"). The remainder of the population, i.e., those who did not develop the disease during the period, is also sampled and similarly classified by presence or absence of the characteristic. Thus, a retrospective study of lung cancer of the same population of 100,000 males over age 40 would (in principle) uncover exactly the same 50 to 75 newly developed cases but would be free to study the characteristics of only a fraction of the remaining 99,925 to 99,950 males who did not develop lung cancer.

Retrospective studies might on the surface appear to supply only estimates of the proportion of persons with and without the disease who possess the characteristic and not to estimate relative risk. Such an estimate can easily be derived, however.[8,9] Denote by P the proportion of the population developing the disease during the time period of interest, i.e., the incidence rate; denote by p_1 the proportion of those developing the disease who possess the characteristic and by p_2 the proportion of those not developing the disease who possess the characteristic. All three proportions can be estimated from a prospective study. Thus, in terms of the notation of Table I:

$$P = \frac{a + c}{N}$$

$$p_1 = \frac{a}{a + c}$$

$$p_2 = \frac{b}{b + d}.$$

The retrospective study will supply estimates of p_1 and p_2 but not of P. The incidence of the disease for those possessing the characteristic is

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

and for those lacking the characteristic

$$\frac{(1 - p_1) P}{(1 - p_1) P + (1 - p_2) (1 - P),}$$

while the relative risk is the ratio of these expressions, namely

$$\frac{p_1}{1 - p_1} \quad \frac{(1 - p_1) P + (1 - p_2) (1 - P)}{p_1 P + p_2 (1 - P)}$$

For investigations in which the retrospective method offers the possibilities of important economies, P will be sufficiently small so that terms containing it may be dropped and the relative risk can be written with only trivial error as

$$\frac{p_1}{1 - p_1} \quad \frac{1 - p_2}{p_2}$$

This ratio depends only on the two proportions p_1 and p_2, estimates of which are supplied by a retrospective study, and not on the over-all incidence, P. Its calculation can be illustrated using the data of Breslow and co-workers,[10] which showed that out of 518 patients with carcinoma of the lung 499 were smokers, while out of 518 controls 462 were smokers. We thus estimate:

$$p_1 = 499/518 = 0.9633$$
$$p_2 = 462/518 = 0.8919$$

$$\text{Relative risk of lung cancer among smokers in terms of unit risk for nonsmokers} = \frac{0.9633}{0.0367} \cdot \frac{0.1081}{0.8919} = 3.2.$$

POTENTIAL SOURCES OF ERROR IN RETROSPECTIVE STUDIES

This somewhat idealized description of the retrospective study sounds so attractive as to make one wonder why any other kind should be considered. In the past, one reason for preferring the prospective study was failure to appreciate that an estimate of relative risk could be supplied by a retrospective study as well. But aside from this reason, which should now be of only historical interest, there are a number of possible sources of error that can arise in the actual conduct of a retrospective study that may take special efforts to eliminate and which may require detailed consideration in appraising the results.[11]

A basic assumption for estimating relative risk from the retrospective study is that it is possible to enumerate all new cases of a disease, or a representative sample of them, without having to observe all the individuals in the population from which they arise and watching for cases to develop. This assumption might be correct if (1) all new patients with the disease sought medical attention, (2) all medical sources to which such patients might go were completely canvassed, and (3) an effective system for reporting such cases was in operation. In practice these conditions may be far from satisfied. Not all new patients seek medical attention, and most investigations confine their canvass to only the most convenient medical source, hospitals. That this may not be sufficient is indicated by the experience of the excellent register of cancer cases maintained by the State of Connecticut. Despite the fact that hospitals reporting to the register cover 94 per cent of the approved general hospital beds in the state, of the 76,000 cancer cases registered during the period 1935 to 1951 almost 19,000 were first discovered by examination of death certificates and consisted largely of patients who received no hospital care or were not reported.[12]

A second and closely related assumption which also requires careful examination is that the sample of individuals not developing the disease supplies an unbiased estimate of the prevalence of the characteristic under study among the entire nondiseased population of interest. Most retrospective studies are content to select a "control" group consisting of individuals with some disease other than that under investigation and to assume that the prevalence of the charac-

teristic in that control group is an unbiased estimate of the required proportion. That this can be a most dangerous assumption is illustrated by Pearl's[13] well-known study of the association between cancer and tuberculosis.

From the first 7,500 autopsies performed at the Johns Hopkins Hospital, Pearl identified 816 individuals with cancer and 816 "control" patients matched for age, sex, race, and date of autopsy. In the control group 16.3 per cent showed active tuberculous lesions, while in the cancer group only 6.6 per cent showed such lesions. A difference in the same direction and of the same magnitude persisted when the material was examined separately for white males, white females, nonwhite males, and nonwhite females. Numerous additional checks on the same material showed the same negative association, or as Pearl called it "antagonism," between the diseases. The possibility that this negative association was causal was investigated by treating terminal cancer patients with tuberculin. We may ask, however, as Wilson[14] did at the time and as Wijsman[15] has more recently, whether this negative association could even be taken as evidence that a population group with active tuberculous lesions at some moment of time would subsequently develop less cancer than a group lacking such lesions, whether or not the association was causal. But if the control autopsy group supplied a biased estimate of the prevalence of active tuberculous lesions among all noncancerous individuals in, say, Baltimore, our answer must be no. A recent examination of Pearl's original records in the Department of Biostatistics of the Johns Hopkins University shows that the control autopsy group included a considerable number of individuals dying from tuberculosis, and therefore, necessarily had a higher prevalence of active tuberculous lesions than that of all live noncancerous individuals. Had the same negative association persisted when the controls consisted of those dying of some disease other than cancer and tuberculosis, it could not have been so easily attributed to a gross sampling bias. But when Carlson and Bell[16] used as a control group only those who died from heart disease, they found the same prevalence of tuberculous lesions as in the cancer group and could not confirm the hypothesis of a negative association. Pearl's result therefore arose not so much from the use of autopsy material as from using it in such a way as to obtain a grossly biased estimate of the prevalence of tuberculous lesions in a live population. Closer attention to representativeness of his controls could have avoided the error.

In Pearl's study the grossly unrepresentative character of the control sample led to an apparently negative association between two diseases. Spurious positive associations can also arise. Thus, if patients in whom a disease occurs in a population do not necessarily seek medical attention, but if an individual with both disease A and disease B is more likely to seek it than one with A alone, this will lead to a spurious positive association.[17]

There are several ways of guarding against the possibility of error arising from the unrepresentative nature of the controls. First of all, controls may be drawn from a wide variety of disease or admission diagnoses. If the prevalence

of the characteristic under study varies widely among the groups, the possibly unrepresentative nature of at least some of them is strongly indicated.[18] It is necessary, of course, that enough controls be studied so that if there are important differences in the prevalence of the characteristic among the different control groups this can be clearly demonstrated. In practice this will often require the study of more controls than of disease cases. A second possible way of proceeding is to draw the control sample from the general population and not from other disease groups available in the hospital. This introduces a possible source of incomparability in the responses of the patients with disease and the controls, since the same question may elicit different answers when asked in radically different situations. A representative sample of incomparable responses does not represent an advance, however, so that the use of general population controls is not necessarily a panacea. The use of both general population and a variety of hospital controls provides a quite general (but not foolproof) safeguard against error from this source.

The reporting in the retrospective study of the presence or absence of the characteristic after the disease status of the respondent is known introduces another potential source of error since conscious or unconscious bias in response may arise. The interviewer who believes, for example, that lung cancer is caused by excessive smoking might be more zealous in questioning a lung cancer patient who gave a nonsmoking history than he would be in questioning a control. A patient's own preconceptions may also influence the answers. The classic example is the report by patients with cancer of the breast of prior injury to the affected breast with considerably greater frequency, and of prior injury to the unaffected breast with considerably less frequency, than was reported by a comparable group of controls.[19] Similarly, in investigations of familial aggregations of disease, families to whom we are led because Susie has the disease may be more likely to remember that grandpa also had it than are families of controls without the disease.

Only empirical investigation can determine whether such memory biases are operating in any particular instance. Thus, Doll and Hill,[20] in their investigation of smoking and lung cancer, interviewed one group of patients who at the time of the interview had been diagnosed as having lung cancer but who were subsequently found to be suffering from some other disease. Their reported smoking habits, however, were similar to those of persons without lung cancer, even though at the time of the interview they had been diagnosed as having the disease. This finding seemed to rule out retrospective reporting error as an explanation of the Doll-Hill results. Clearly, whenever double-blind interviewing is possible, it will control this source of error.

This catalogue of potential sources of error is not intended to be, neither should it be construed as, a blanket condemnation of the retrospective method or of any particular set of study findings yielded by it. The magnitude of error in any particular case is a substantive issue to be resolved on its own merits.

Collateral evidence can provide information on possible magnitudes of different errors and the size of the spurious association that could result. Sweeping condemnation of the retrospective method or uncritical acceptance of the results of single studies are equally to be avoided. The frame of mind which condemns any method that could lead to error under some conceivable set of circumstances, without also considering whether those circumstances have in fact arisen, is unlikely to be satisfied with any result outside the field of pure mathematics. The contrary frame of mind, which accepts a method simply because it will yield an answer without consideration of how much in error the answer could be, is scarcely likely to be any more productive. The retrospective study provides an economical but not a foolproof method of studying certain types of relations. Its results, like all results in science, must be checked in a variety of other ways before they can be accepted with confidence.

MEASURES OF ASSOCIATION

Why take the ratio of the incidence rates of those with and without the characteristic as a measure of association? Why not some other combination of the two rates, such as the difference? There is no general agreement on the answers to these questions[5,6] and we can do no more than present our own point of view.[21]

It is well to begin by recognizing that no single combination of the two incidence rates contains all the information yielded by the two separate rates. Given only the difference or the ratio or some other single combination, it is not possible to reconstruct the individual rates on which the original combination is based. The idea that there is necessarily a single measure which uniquely and comprehensively summarizes all aspects of an association seems erroneous. To talk, therefore, about *the* measure of the causal effect of an agent is to talk about a hypothetic phenomenon whose existence remains to be demonstrated. We would suggest that the appropriateness of any measure of association is to be judged by whether it helps in the understanding of the phenomena under study and not by any formal mathematical criterion of rightness or wrongness.

From this point of view, both the ratio and the difference of incidence rates serve a purpose. Thus, if one accepts any observational associations found as causal and inquires about their importance, a natural measure to use is the excess number of cases of the disease attributable to the characteristic. This is equivalent to taking the difference in incidence rates as a measure of association. If, however, the existence of an observational association cannot automatically be taken to imply a causal association as well, and this is the usual situation, other measures are required. An observational association may merely reflect the presence of some other, unknown common cause. It is sometimes suggested, for example, that cigarette smoke is not really a causal agent in the development of lung cancer but that some special constitutional make-up, perhaps genetic in origin, predisposes certain individuals to lung cancer and also makes them

cigarette smokers. It is in the investigation of such possibilities that the ratio of incidence rates is most useful. Thus, cigarette smokers have been reported to have a ninefold greater risk of dying of lung cancer than nonsmokers. If this observational association is not causal but merely reflects the common effect of some unknown third characteristic, one can say that this third characteristic must be at least ninefold more prevalent among cigarette smokers than among nonsmokers.[21] If quantitative investigation shows that the relative prevalence (cigarette smokers to nonsmokers) of suspected common characteristics is less than ninefold, these characteristics cannot by themselves account for the observed association. On the other hand, if one is told that the difference in annual incidence or mortality rates between cigarette smokers and nonsmokers is 40 per 100,000, nothing about the difference in the prevalence of the postulated third characteristic between cigarette smokers and nonsmokers can be inferred.

Similarly, if a single agent is associated with two diseases, we may say that the association with the disease having the higher relative risk is less likely to be explained by a common third cause. Thus, the 70 per cent elevation in risk from coronary heart disease among cigarette smokers that has been reported could possibly be explained as the result of a common third characteristic whose prevalence among cigarette smokers is twice that among nonsmokers. It would be arithmetically impossible, however, for this same characteristic to explain the ninefold difference in lung cancer.

A second useful purpose served by the relative, but not the absolute, measure is in the refinement of classification. A priori considerations will not often indicate exactly what the classification by characteristic should be; neither will they indicate exactly how the disease should be defined. Thus, smokers may be defined as those who smoke either cigarettes, cigars, or pipes or as those who smoke only cigarettes. Lung cancer may be defined to include all histologic types or restricted to epidermoid carcinoma of the lung. There is a precise sense[21] in which one can say that the best classification on both the characteristic and the disease axes is the one leading to the largest relative risk. Thus, the stronger association of cigarette smoking with epidermoid carcinoma of the lung than with adenocarcinoma, as reflected in a larger relative risk for the former, suggests that adenocarcinoma may not be related to smoking. Use of the differences rather than the ratios of incidence rates would not reveal this.

Finally, the incidence ratio provides some indication of the importance of characteristics other than the one being studied. If the characteristic under study is only one of many independent characteristics associated with the disease, the ratio will be closer to unity than if this is not the case. The presence of many other possible causes, furthermore, indicates the necessity of exercising great caution in attributing causal significance to the observed association for any one characteristic.

It has been suggested[5,6] that when other causes are present it is incorrect, or at least inappropriate, to measure effects using the ratio of observed incidence rates, since these are compounds of more fundamental rates. But the rates postulated in competitive risk models are themselves compounds of more fundamental constants, such as reaction velocities, which could in their turn be deduced from even more fundamental physical considerations. The process of expanding scientific understanding is a never-ending one, and all that can reasonably be asked of a descriptive or analytic method is that it contribute to this process.

ELIMINATING THE EFFECTS OF OTHER VARIABLES

The discussion up to this point has assumed that the measures of association can be computed without regard to the effects of other known or suspected variables. This assumption will rarely be true, however, and methods of eliminating the possible effects of other variables must be considered. If extraneous variables are very highly correlated with the characteristic under study, or if they are too ill-defined to be measurable, of course little can be done to eliminate their effects in observational studies. Two general methods are available for eliminating the effects of variables that do not fall in this category. The first involves matching disease and control cases with respect to the control variables. The second calls for the selection of independent, unmatched samples of cases of disease and controls, but with the effects of extraneous variables eliminated in the subsequent statistical analysis. It is possible to combine both procedures by matching on some variables (normally, those whose importance has been previously established) and analyzing for the effects of others.

The most general procedure of analysis involves cross classification of each of the two samples with respect to the extraneous variables and separate (at least in principle) computation of the association between the characteristic of interest and disease status for each basic cell of the cross classification. Extensive cross classifications require an abundance of observational material. In such cases it is natural to group quantitative variables into broad classes. The loss attributable to overly coarse grouping does not appear to have been studied, although results of Cox[22] on a different but related problem indicate that as few as two classes of equal size will retain two-thirds of the information present, while four equal-size ones will retain 86 per cent. The analysis of covariance as an alternative to cross classification is sometimes employed,[23] although usually there are restrictive assumptions made as to the absence of interactions and the presence of only linear effects. Postmatching is a device by which the samples of cases of disease and controls, originally selected independently, are matched at the conclusion of field work; data for unmatched subjects in both groups are discarded. In addition to its obvious lack of efficiency, this method has the further disadvantage of making the study of interactions difficult.

Prematching, although much used, is usually difficult to carry out in the field because controls which match on a series of variables are not easy to obtain.

As in postmatching studies, if interactions among variables exist, the interpretation of results may be equivocal. Furthermore, the effects of the matched variables themselves on the disease cannot be studied.

The analysis of tables of multiple classification is too large a subject to enter into here. One aspect, however, the estimation of relative risk, requires mention. For each cell of a multiple classification one obtains an estimate of relative risk, say r. Some method of combining the estimates for the different cells is required. A combined estimate in the form of a weighted average seems reasonable, but different criteria in choosing weights can lead to different results. Mantel and Haenszel[18] propose as a compromise weight for a cell

$$\frac{p_2 (1 - p_1)}{1/n_1 + 1/n_2}$$

where p_1 and p_2 are (as explained previously) the proportions with the characteristics in disease and control groups in that cell and n_1 and n_2 are total number of cases of disease and controls studied in that cell. This is an easy estimate to compute since it reduces, in the notation of Table I, to

$$\text{Combined relative risk} = \Sigma \frac{ad}{N} \Big/ \Sigma \frac{bc}{N}.$$

It is particularly important to note that common methods of combining cells, such as pooling, or analogues of the direct or indirect method of age-adjustment[24] yield results that need not be weighted averages of the individual relative risks and can, therefore, give an over-all relative risk which is entirely outside the range of the relative risks observed for each cell. As an extreme example of what could conceivably happen, Mantel and Haenszel give the following example. Consider two cells, in the first of which $p_1 = 0.05$ and $p_2 = 0.01$ and in the second of which $p_1 = 0.99$ and $p_2 = 0.95$. Each cell thus gives a relative risk of approximately 5. If equal numbers had been covered in each cell and the two cells were combined by pooling, however, the relative risk of the combined cell becomes not 5, but 1.3. Although such an extreme result might not often be encountered in practice, methods of estimation that can lead to it cannot be generally recommended.

The warning as to the potentially misleading nature of the estimate of relative risk yielded by pooling applies to matched-sample studies as well. These may be considered as a special case of multiple classification with the number of cells studied equal to the number of pairs included. In such a study any pair can yield one of four possible results: (1) a disease case with the characteristic, control case without it, (2) disease case without, control case with, (3) both with, and (4) both without. Call the number of pairs of the first type a, of the second type b, of the third type c, and of the fourth type d. Kraus,[25] in a consideration of this problem, recommended as an estimate of relative risk a/b. Interestingly enough, the weighted risk proposed by Mantel and Haenszel for the general multiple classification case reduces to a/b for the matched sample case, as does

the maximum likelihood estimate of the relative risk on the assumption that all pairs have the same relative risk. The estimate which would result from pooling the data in all four categories, $\frac{(a+c)\,(a+d)}{(b+d)\,(b+c)}$, is different, however, and not easy to defend.

REFERENCES

1. Goldberger, J., Waring, C. H., and Tenner, W. F.: Pellagra Prevention by Diet Among Institutional Inmates, Pub. Health Rep. **38**:2361, 1923.
2. Dorn, H. F.: Methods of Measuring Incidence and Prevalence of Disease, Am. J. Pub. Health **41**:271, 1951.
3. Neyman, J.: Statistics—Servant of All Sciences, Science **122**:401, 1955.
4. Sartwell, P. E., and Merrell, M.: Influence of the Dynamic Character of Chronic Disease on the Interpretation of Morbidity Rates, Am. J. Pub. Health. **42**:579, 1952.
5. Berkson, J.: The Statistical Investigation of Smoking and Cancer of the Lung, Staff Meet. Mayo Clin. **34**:206, 1959.
6. Sheps, M. C.: An Examination of Some Methods of Comparing Several Rates or Proportions, Biometrics **15**:87, 1959.
7. Goodman, L. A., and Kruskal, W. H.: Measures of Associations for Cross Classifications, J. Am. Stat. A. **49**:732, 1954.
8. Cornfield, J.: A Method of Estimating Comparative Rates From Clinical Data, Applications to Cancer of the Lung, Breast and Cervix, J. Nat. Cancer Inst. **11**:1269, 1951.
9. Cornfield, J.: A Statistical Problem Arising From Retrospective Studies, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1956, University of California Press, vol. 4, pp. 135-148.
10. Breslow, L., Hoaglin, L., Rasmussen, G., and Abrams, H. K.: Occupations and Cigarette Smoking as Factors in Lung Cancer, Am. J. Pub. Health **44**:171, 1954.
11. Dorn, H. F.: Some Problems Arising in Prospective and Retrospective Studies of the Etiology of Disease, New England J. Med. **261**:571, 1959.
12. Griswold, M. H., and Cutler, S. J.: The Connecticut Cancer Register, Seventeen Years of Experience, Connecticut M. J. **20**:366, 1956.
13. Pearl, R.: Cancer and Tuberculosis, Am. J. Hyg. **9**:97, 1929.
14. Wilson, E. B.: Morbidity and the Association of Morbid Conditions, J. Prev. Med. **4**:27, 1930.
15. Wijsman, R. A.: Contribution to the Study of the Question of Association Between Two Diseases, Human Biol. **30**:219, 1958.
16. Carlson, H. A., and Bell, E. T.: A Statistical Study of the Occurrence of Cancer and Tuberculosis in 11,195 Post-Mortem Examinations, J. Cancer Res. **13**:126, 1929.
17. Berkson, J.: Limitations of the Application of Fourfold Table Analysis to Hospital Data, Biometrics **2**:47, 1946.
18. Mantel, N., and Haenszel, W.: Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease, J. Nat. Cancer Inst. **22**:719, 1959.
19. Lane-Claypon, J. E.: A Further Report on Cancer of the Breast, With Special Reference to Its Associated Antecedent Conditions, Rep. Pub. Health & M. Subj. No. 32, pp. 1-189, 1926.
20. Doll, R., and Hill, A. B.: A Study of the Aetiology of Carcinoma of the Lung, Brit. M. J. **2**:1271, 1952.
21. Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L.: Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions, J. Nat. Cancer Inst. **22**:173, 1959.
22. Cox, D. R.: Note on Grouping, J. Am. Stat. A. **52**:543, 1957.
23. Wells, H. B., Greenberg, B. G., and Donnelly, J. F.: North Carolina Fetal and Neonatal Death Study. 1. Study Design and Some Preliminary Results, Am. J. Pub. Health **48**:1583, 1958.
24. Woolsey, T. D.: Adjusted Death Rates and Other Indices of Mortality, *in* U. S. Bureau of the Census: Vital Statistics Rates in the United States, 1900-1940, Washington, D. C., 1943, U. S. Government Printing Office, ch. IV.
25. Kraus, A. S.: The Use of Family Members as Controls in the Study of the Possible Etiologic Factors of a Disease, Dissertation in Partial Fulfillment of the Requirements for the Sc.D., Department of Biostatistics, School of Public Health, University of Pittsburgh, 1958.