

Elucidation of some epidemiologic principles

by Olav Axelsson, MD¹

Axelsson O. Elucidation of some epidemiologic principles. *Scand j work environ health* 9 (1983) 231–240. This review is an attempt to elucidate some fundamental epidemiologic concepts, particularly various aspects of the case-referent design and its relations to the cohort type of study. Some of these aspects also have direct implications for judgements about the comparability of differently exposed population groups, eg, with regard to confounding.

Key terms: Case-referent (case-control), cohort, confounding, effect modification, interaction, study base, study population.

During the past decade epidemiologic research has become an important area in occupational health, just as it has in many other fields of medicine. The epidemiologic concepts and methods have developed, and the resulting principles and viewpoints are of primary interest not only for those engaged in research, but also for others having to interpret and evaluate the findings that are continuously being presented. This paper is an attempt to facilitate the understanding of some basic epidemiologic concepts and viewpoints in recent terms, particularly regarding the relationship between case-referent and cohort studies. Some aspects of the interference between the various factors that determine the occurrence of disease in a population are also considered. Such factors might be generally referred to as determinants of the occurrence of disease, but they do not necessarily imply causal relationships. The term risk indicator is also commonly used and would serve synonymously, but preferably in the context of increased morbidity.

On the relationship between case-referent and cohort studies

A consideration of epidemiologic study designs may start from a distinction between the various absolute measures of occurrence of disease, namely, (i) incidence rate or incidence density, ie, the number of incident cases (or deaths) per person-years at observation, (ii) cumulative incidence or cumulative risk, ie, the fraction of individuals falling ill (or dying) during a defined period of time, and (iii) prevalence rate, ie, the fraction of individuals being ill at a certain point in time. These measures will be more thoroughly dealt with later on, but it might already be recalled that such rates can be compared for different (sub)populations in terms of relative measures like rate ratios or rate differences. Furthermore each one of these rates has its characteristic counterpart with regard to the type of study (6) that can be applied in epidemiologic research. In a discussion of these various study designs it is reasonable to start from the commonly used and easily understood cohort approach, in relation to which the specific features of the conceptually more difficult case-referent design can then be more thoroughly delineated. Before the details of these aspects are gone into, however, the recently introduced concept of study base (7) might be considered as

¹ Department of Occupational Medicine, University Hospital, Linköping, Sweden.

Reprint requests to: Prof O Axelsson, Department of Occupational Medicine, University Hospital, S-581 85 Linköping, Sweden.

helpful for the understanding of study designs.

Any particular epidemiologic study derives its information from an observation of a population for a longer or shorter period of time, the cross-sectional or prevalence study being a special case with a momentary time span. The specific population under study can be a larger or smaller proportion of the general population of a country, a country, or some other administrative unit, and it might be referred to as the study population. If representatively chosen, a random sample, say, its experience over time, would constitute a base for descriptive information about the occurrence of disease in the general, or source, population. Synonymously with study population, the term base population might also be attractive, ie, when one merely thinks of the individuals themselves rather than of their experience over time, as included in the base concept.

For etiologic or interventive research, ie, hypothesis-testing epidemiology focusing on causal relationships, the base population is usually chosen to encompass a specific segment of the general population. Then, the determinant at issue, which is usually an earlier and/or a current exposure to an industrial chemical or some other factor, should be present among a suitable number of individuals and absent for others. However, it may be even more preferable to enroll individuals with varying degrees of exposure so as to obtain different exposure categories in the base population for the purpose of also studying dose-response relationships. Ideally, the nonexposed segment of the base should be selected to become as similar as possible to the exposed one(s) in terms of natural or background morbidity and with regard to the possibility of obtaining accurate information about the exposure(s) and disease(s) under study. Considerations about comparability of the exposed and nonexposed one or more segments of the base might be referred to as an issue of so-called validity, which will be dealt with to some extent in this overview, and further discussions are available elsewhere (1, 2, 4, 5, 9).

Any general population has a turnover in membership because of migration, death, and birth, ie, it is open with regard to its

members, and similarly a base population can also be open in character. Due to the turnover of members in an open population, the measures or occurrence of disease, either absolute or relative, have to relate the cases to a denominator of person-years at observation (with the risk of falling ill) rather than to the individuals, since they are continuously exchanged. Hence the incidence rate (incidence density) is the appropriate measure of occurrence of disease in an open population.

On the other hand, if the base population is closed, ie, if the individuals have been primarily defined and enrolled to constitute the base, then the measures of occurrence of disease may have individuals in the denominator, the result being the cumulative incidence or, alternatively, the prevalence rate if the time span of the study is momentary; the corresponding rate ratios or differences are obtained accordingly. Given a time extension of the study, a person-year denominator is still a possibility however, and it is utilized in certain designs. Hence the expected numbers of deaths or cancers might be calculated in a closed population for a particular time period by the multiplying of the person-years at observation by the corresponding national age- and gender-specific rates. A comparison of this sort between a particularly selected base population and the general population is always questionable however, an aspect that has been heavily stressed quite recently (9).

Irrespective of whether the base population is open or closed, it might be helpful to view the base as shown in fig 1, where the cases are symbolized with the plus signs and the total area might be taken as representing the size of the base over time, ie, the background of the cases either in terms of person-years or number of individuals. An open base should be thought of as "emitting" its cases, since the case individuals no longer provide person-years at observation with regard to risk of falling ill. To the contrary, a closed base retains the cases since the applied measures of disease occurrence have count denominators, ie, all the individuals originally enrolled, and therefore also include the cases. If a closed base popula-

tion is viewed in terms of person-years of follow-up, however (cf preceding discussion), again the cases are "emitted" from the active, case-producing part of the base, since an individual is not providing any more person-years at risk after the inception of the disease at issue or after death, whichever outcome is being studied.

As shown in the table of fig 1, the disease rates may be obtained for the different domains of the determinant (a "two-point design" being assumed) with knowledge of the number of cases and the sizes of the respective domains in terms of denominators, here taken as proportional to the area units, or base units, as indicated in the figure. If all the individuals or person-years of the two domains are known and a comparison is made through the ratio or the difference of the respective rates, one would have the traditional follow-up or cohort approach. However, the rate ratio could be calculated even if only the relative sizes of the exposed and nonexposed domains in the base were known. In this respect the necessary information would be obtainable already through a sampling of the base, ie, through the selection of referents likely to be proportional in number to the base units in the respective domains of the figure. Such sampling is well-known under the concept of a case-referent (case-control) study. Hence, as illustrated, there is obviously an inherent and close relationship between cohort and case-referent studies, although they are often thought of as representing different principles in epidemiologic designs.

Study designs based on incidence rate (or incidence density)

When a study is characterized by an open base population, the new cases of the illness at issue, which occur in the base during the study period, would be either exposed, *a*, or nonexposed, *b*. As already stated, the cohort approach means a comparison of the incidence rates among exposed and nonexposed subjects and would require the calculation of the person-years in the denominators. However relative information about the person-years in the base may be obtained through a direct sampling of the base population over time. The desired information might also be achieved in a somewhat more indirect way, namely, by taking individuals with noncase diagnoses to represent the population over time and, hence, also the distribution of person-years over categories of the determinant, say, a particular chemical exposure. If these referent individuals are denoted by *c* when exposed and by *d* when nonexposed and the study period is *t'*-*t''*, the incidence rate ratio (or incidence density ratio) may be constructed as

$$IR_1/IR_0 = \{a/[kc(t''-t')]\}/\{b/[kd(t''-t')]\}, \text{ (eq 1)}$$

where the incidence rate among the exposed is denoted by *IR*₁ and *IR*₀ is the incidence rate among the nonexposed; the constant *k* is the inverted value of the (unknown) fraction of the base population as sampled through the case-referent approach. Consequently the denominators represent the exposed and nonexposed

	A-	A+
	+	+
	+	+
	+	+
	+	+
	+	+
	+	+
	+	+

	A+	A-	RR
Cases	12	6	4
Base (units)	3u	6u	
Rate	12/3u	6/6u	

Fig 1. An illustration of the presence and absence of a determinant, A, in the study base. The base might be represented either by a sample or by a census; u denotes units of the base. Hence the rate ratio $RR = (12/3u)/(6/6u) = 4$.

person-years of the base, from which the exposed and nonexposed cases, a and b , respectively, are "emitted." If the size of the sample is finally permitted to increase to encompass all individuals in the base for each year, ie, a census, then the study obtains the character of a cohort study with incidence rates (having person-year denominators).

With the case-referent view, the abbreviations in the preceding relation (equation 1) result in

$$IR_1/IR_0 = ad/(bc), \quad (\text{eq 2})$$

and thus the incidence rate ratio is calculable in a simple and well-known way if the exposure status is known for the cases and a sample of the study population, ie, the referents, even though the actual number of person-years at risk in the base remains unknown. Furthermore all cases do not necessarily have to be enrolled in the study, but the proportions of exposed and nonexposed cases in the sample need to be the same as in the base population (6).

Since c and d are supposed to reflect the exposure situation for the healthy individuals in the base (healthy with regard to the disease under study) throughout the study period, then, in principle, a noncase individual enrolled as a referent in the earlier part of the study period might later become a case and thus be represented twice in the same study. If deceased referents (being noncases with regard to diagnoses) are used, the principle of the incidence rate type of case-referent study would still work, since dead noncases (also having been "emitted" from the base population) apparently reflect the exposure situation of the base population over the study period. One avoids, however, the intellectual frustration of having the same individual primarily enrolled as a referent and later perhaps developing the disease under study and then being considered a case. In viewing (ie, sampling) the base through cases of some other disease entity than that under study, it is apparently important that this disorder has no relation (either positive or negative) to the determinant at issue, either directly or indirectly. Should a particular exposure cause some other disease than that under study, this other disorder cannot be

utilized as a reference entity since it would not reflect the distribution of the exposure in the base properly. If hospital registers are used, a similar referral pattern should result (eg, with regard to social and geographic background) both for the disease under study and for the other disorder, which may constitute the referent series, since a hospital register is not always a fully representative source of information about all the cases occurring in a population. The same applies if several different disease entities are utilized for the referent series.

The type of case-referent study which has just been described is the common one in epidemiologic research, but its specific structure has not been generally recognized as related to an open population with dead or "immune" individuals being continuously replaced. (In this context "immune" is taken in a broad sense; eg, technically speaking, an individual surviving a heart attack might be considered "immune" to that particular type of disease.) Nor has it been particularly emphasized that the referents (or "controls") are there to provide (relative) information about the relation of exposed and nonexposed person-years in the base. Finally, the rate ratio, as derived from a case-referent study of this type, is commonly referred to as an (exposure) "odds ratio," but there remains little justification for this concept with regard to its derivation under the delineated principles.

Study designs based on cumulative incidence

If cases are accumulated over time from a closed or static population, count denominators would be appropriate for the comparison of the morbidity or mortality among exposed and nonexposed individuals. These denominators would simply be the number of individuals originally enrolled in the exposed and nonexposed domains of the base population. Now, if instead a sample is taken from among those of the base population who remain healthy at the end of the study period, one obtains the cumulative incidence type of case-referent study, which could be thought of as nested within a cohort, ie, in a closed population without

any turnover through replacement of dead or "immune" individuals.

Again let a and b represent exposed and nonexposed cases originating from the exposed and nonexposed subdomains of the base population, M_1 and M_0 , respectively. Now, retaining the cases and having no turnover in the base population, the source of referents (taken as noncases) obtains the structure $M_1 - a$ and $M_0 - b$, respectively, at the end of the study period. These quantities approximate M_1 and M_0 only so far as the disease is rare, namely, if a and b are small fractions of M_1 and M_0 . The risk ratio or the ratio of the cumulative incidences, CI_1 and CI_0 , might be taken as

$$CI_1/CI_0 = (a/M_1)/(b/M_0) \doteq [a/(M_1 - a)]/[b/(M_0 - b)] \quad (\text{eq 3})$$

or

$$CI_1/CI_0 \doteq (a/b)/[(M_0 - b)/(M_1 - a)] = ad/bc \quad (\text{eq 4})$$

since the quotient $(M_1 - a)/(M_0 - b)$ is represented by the quotient of exposed to nonexposed referents, c/d .

This type of rate ratio might well be referred to as an "odds ratio" to indicate its character (and approximateness) since, as already indicated, this sort of case-referent sampling requires that the disease under study be rare to provide a reasonably correct approximation of the risk ratio. Although it has been traditional to assume a rare disease condition for the "case-referent" design to be applicable, nevertheless the incidence rate type of study has usually been applied rather than this "classical" rare disease type.

A variant of the cumulative incidence type of study can also be obtained without involving the rare disease condition, namely, if the referents, c and d , are chosen as a sample from the total base population rather than from its noncase domain, ie, some cases would be permitted to appear in c and d if they happen to be caught in the sampling of the base. However, there might not always be suitable registers available for this modified approach of the cumulative incidence study, which may explain why it has been rarely applied in spite of its theoretical advantages; compare also the corresponding aspect in the context of studies with prevalence rate as discussed later.

Study designs based on prevalence rate

By considering the base population at a particular point in time rather than over a period, one obtains a cross-sectional study, and the measure of disease is prevalence rate and its relative derivatives. Again there is a choice between just a sample of the base population or a census approach. Should a sampling of the base be preferred, ie, the case-referent approach, there are two further options with regard to the acquisition of referents. First, if the referents are taken as noncase individuals, which might be felt as the "natural" thing to do in a case-referent situation, one will finally arrive at the incidence rate (or incidence density) ratio and will not obtain the prevalence rate ratio, as perhaps would be expected. It is required for an unbiased estimate, however, that the duration of the disease not be influenced by the exposure, which is probably the usual situation. These viewpoints are further elaborated in the appendix.

The other option in selecting referents would be to take a cross-sectional sample, T_p , of individuals out of the base population, irrespective of whether or not they are healthy with regard to the disease under study. This sample would provide the ratio of exposed to nonexposed individuals. Then the ratio of the prevalence rates (by definition the prevalence rate contains both cases and noncases in the denominator) is obtained when the quotient of exposed to nonexposed cases is multiplied by the ratio of nonexposed to exposed individuals in the base population sample (d_p and c_p , respectively, being the components of T_p). With the usual denotation one obtains the prevalence rate ratio (PRR) as

$$PRR = (a/b)(d_p/c_p).$$

Now, recall that d_p and c_p may include some of the cases which compose a and b , since the sample T_p was drawn from the base population regardless of the individuals' health status. Apparently this approach can be an efficient and economic variant of the cross-sectional study, since not every individual has to be dealt with in the data acquisition, eg, in the checking of various background parameters like age, smoking, domicile, etc. Furthermore, if the sampling fraction is known, then also the

prevalence rates can be obtained after the multiplication of c_p and d_p by a corresponding factor (ie, by 20 for a 5 % sample, 10 for a 10 % sample, 5 for a 20 % sample, etc). Notice finally that, if the disease is rare, the incidence rate ratio comes close to the prevalence rate ratio, which is also reflected by the fact that individuals with a rare disease are unlikely to be caught in a sample from the base population, ie, to be included in c_p or d_p .

Reporting of the study design

Usually there is no reporting in epidemiologic papers about how the base of a particular study was defined or how it should be thought of. Nevertheless, for the so-called cohort studies, the base population is usually in clear sight as the particular population described in the study, but in case-referent approaches the base is less explicitly accounted for but merely indirectly referred to through the definition of the period of time for the study and the way of ascertaining the cases. Sometimes the base might have "scotomas" or holes, however, ie, for social or ethnical reasons – say, parts of the potential study population might not deliver cases to the particular register or hospital used for the study. Such "scotomas" should then also be accounted for in the base sampling through the acquisition of the referents so as to achieve comparability. There are apparently great difficulties involved in judging whether such phenomena take place or not, and consequently referents taken as noncase disorders from the same register or hospital, and with a similar referral pattern as the cases, may be preferable to a population sample if a situation of this sort is suspected. On the other hand there might be difficulties in judging which specific noncase diagnoses would be suitable referents as without any relation to the exposure. A base population, which is intended for a cohort approach, is of course also feasible for subsequent application of the case-referent technique. Then the efforts of the study are reduced to the collection of cases and a limited number of referents, an approach which is particularly economical if extensive medical examinations have to be carried out.

In the reporting of results of epidemiologic research there is normally little distinction made with regard to the specific type of the case-referent studies that have been carried out. This is also of little practical importance so long as the disease is relatively rare, which is usually the situation. In the context of frequent disorders however, eg, cardiovascular disease, bronchitis or certain common symptoms, etc, one has to be more careful in the design of the case-referent study. In general, the incidence rate type of study might be preferred as providing good estimates of the epidemiologic parameters. Sometimes mosaic types are created, and, again, they are acceptable if the disease is rare. Nevertheless a distinction between the various types of the case-referent sampling approach is of importance for a better understanding of the principles behind this technique and the relations to cohort designs and the various measures of disease, but also for the understanding of different viewpoints with regard to validity.

Some remarks on confounding, interaction and effect modification

Rarely, if ever, does only one factor determine the occurrence of a disease in the population, and therefore one has to consider several factors with greater or lesser influence. Quite often various determinants also tend to be aggregated with each other so that difficulties arise in deciding the actual impact from either of them, which means confounding. More precisely, if an already known risk indicator is associated with the potential determinant under study, confounding is present (4). This situation should not be confused with interaction however, ie, when one or both of two factors tend to influence the effect of the other, or of each other, in the causation of a particular disorder (8). Complicated overall patterns might arise since both interaction and confounding may take place concomitantly.

Fig 2 shows a simple situation in the base, namely, the presence of two independently operating determinants, A and B, and without any interaction between them, ie, the model is additive (cf later

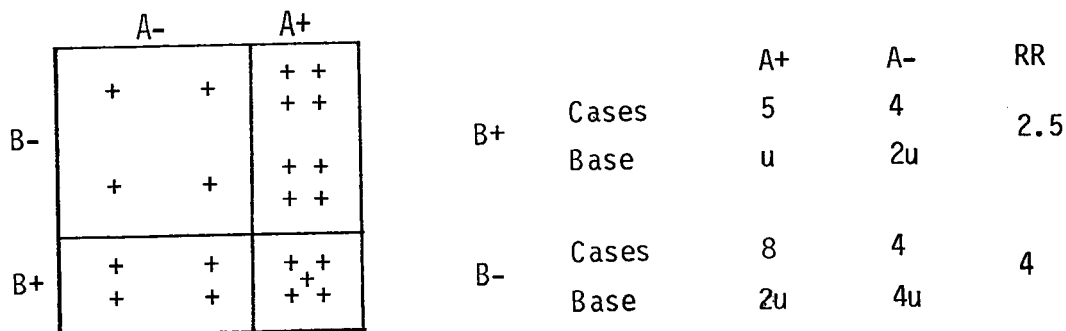


Fig 2. Study base with two independently operating determinants, A and B, ie, an additive model.

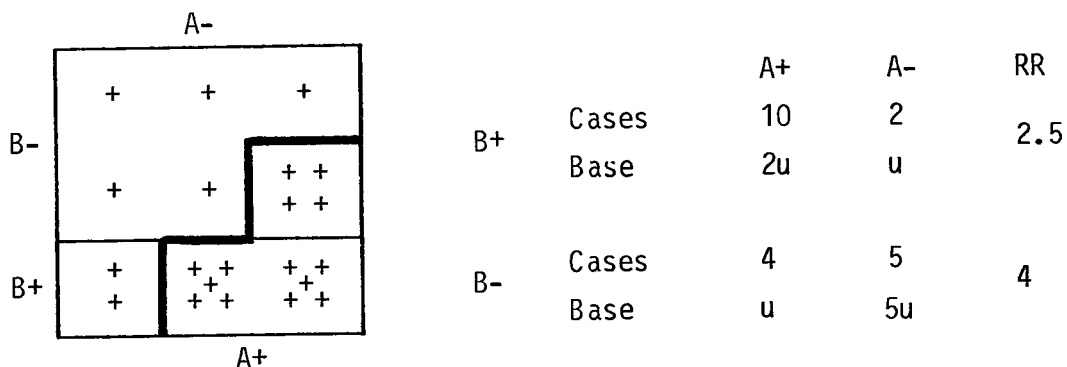


Fig 3. Study base with confounding between two independently operating determinants, as in the previous figure. Note relationship between A+ and A- in the B+/B- sections of the base however, ie, $2u/u = u/5u$ (confounding).

comments with regard to competition, however).

Regardless of whether the study is thought of as a cohort or a case-referent study, the data obtained from the base might be presented as in the table of fig 2. Note that the ratio of A (exposed) to non-A (nonexposed) in the base is the same, both in the presence and in the absence of B, as reflected by the base (sample) in the table, and hence there is no association between the two determinants and therefore no confounding. If the effect of A is considered, the stratum-specific rate ratios might be calculated, and one obtains 4.0 in the absence of B and 2.5 in its presence. This difference is obviously due to the fact that the background occurrence of disease in the B domain is increased through the effect of B, and therefore the magnitude of the effect from A is relatively less. Thus effect modification is present in terms of the rate ratio (whereas the rate difference is the same in both strata), but still there is no interaction between the factors in this

additive model. Should both A and B have a very strong and similar effect or exert mutually exclusive effects on each other, there would be a competitive interaction however; compare the phenomenon of competing causes of death.

Fig 3 shows an additive situation similar to that in fig 2, but now with some association between A and B, ie, confounding is at hand, as can be seen directly in the figure and also in the table in a comparison of the ratio of the exposed to the non-exposed in the base representation of the two strata. It is worth considering in this context that the sampling of the base in a case-referent approach would be influenced by random variation, and a confounded situation in the base would therefore not necessarily be reflected in the data - nor does some confounding in the case-referent data themselves necessarily mean the presence of confounding in the base.

In fig 4 a base is illustrated in which both interaction and confounding is

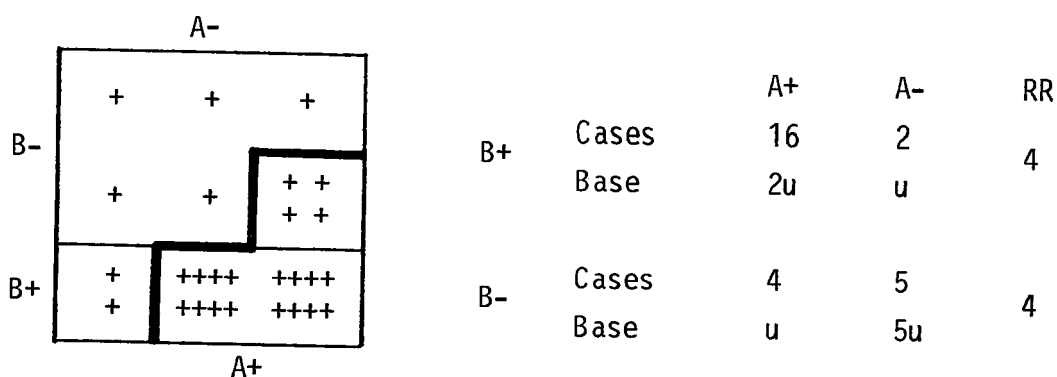


Fig 4. Study base with confounding and multiplicative interaction between the two determinants A and B. The rate ratio for A+ and B+ combined is obtained as $RR_{AB} = (16/2u)/(5/5u) = 8$ or in "case-referent thinking" $RR_{AB} = (16 \times 5u)/(5 \times 2u) = 8$.

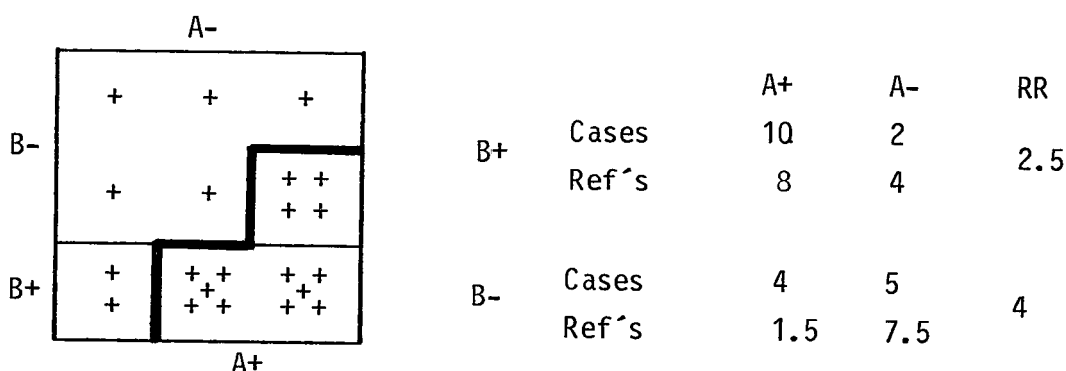


Fig 5. Effect of sampling the base from fig 3 in a case-referent approach and dissolving the pairs. Note that by matching on B with one referent per case the probability of obtaining an A + B + referent is $(10 \times 2/3) + (2 \times 2/3) = 8$ etc, $2/3$ representing the A + B + domain of the total B + sector. (Ref's = referents)

present. The model is now constructed as multiplicative and reflects a situation which is often more or less clearly present in practice. The rate ratios for A as a determinant are now similar both in the presence and in the absence of B, ie, there is no effect modification in terms of the rate ratio, which remains constant in the presence of a multiplicative interaction. Instead there is effect modification in terms of the rate difference. This matter may perhaps appear contradictory at first but is inherent in the utilization of the rate ratio as a relative measure of the occurrence of disorder. It should be noted also that the combined effect of A and B is obtained if one takes the rate ratio between that subdomain of the base, where both A and B are present, and the subdomain, where none of the factors operate (ie, going diagonally between the strata in the tables of the respective figures).

Finally, it might be of interest to use fig 3 for determining the consequences of matching on the confounder B for studying the effect of A, ie, each time a case appears in the B domain; also the referent should be chosen from this sector of the base. The result of this matching procedure appears in fig 5, where the cases and referents, as obtained through matching, are presented in the respective strata according to what would be the expected distribution on the mean (therefore decimals, ie, the proportion is 1.5 versus 7.5 referents in the non-B stratum). The rate ratio is unchanged in the respective strata, but the actual distribution of A and B in the base is no longer reflected in the referent sample, ie, the A+, B+ domain appears as quite large, as does A-, B+, whereas the A+, B- and A-, B+ are indicated as relatively smaller. This distorted base sample would not be

suitable for an evaluation of the effect of some other determinant that might turn out to be of interest during the course of the study, namely, if this new determinant of interest has an uneven distribution in the base; see however chapter 7 of the communication by Breslow & Day (3) for methods coping with this sort of problem.

From the preceding example it can also be seen that matching on a risk indicator tends to increase the efficiency of a study, since matching leads to an increase of the number of individuals in the B domain. In this sector the effect of A is relatively weaker in relation to the reference level of morbidity, which is elevated through the influence of B. Therefore, in the absence of matching, the effect in this stratum would be less likely to appear as "significant" by statistical testing in a study of a given size.

Closing remarks

The design phase of an epidemiologic investigation is usually crucial for the success of the intended study. In this paper an attempt has been made to elucidate some of the fundamental aspects and concepts that may be helpful for the understanding of epidemiologic research. Especially the close relationships of the cohort and case-referent designs become quite apparent through the newly introduced idea of thinking in terms of a study base. It is also felt that this concept is very valuable in considerations of various epidemiologic phenomena, eg, issues of interaction and confounding. However, none of the aspects brought forward in this paper are actually new, but the way

of viewing some of the basic epidemiologic principles is slightly different and somewhat refined and thereby hopefully also sheds light on some of the more or less difficult matters involved in epidemiologic research.

Acknowledgment

I am indebted to Prof OS Miettinen for the many discussions, over the years, that underlay this paper and to the Institute of Occupational Health in Helsinki for the invitation to attend the seminar on epidemiologic methods in 1982.

References

1. Axelson O. Aspects on confounding in occupational health epidemiology. *Scand j work environ health* 4 (1978) 85–89.
2. Axelson O. The case-referent (case-control) study in occupational health epidemiology. *Scand j work environ health* 5 (1979) 91–99.
3. Breslow NE, Day NE. Statistical methods in cancer research. Volume 1. International Agency for Research on Cancer, Lyon 1980.
4. Hernberg S. Epidemiology in occupational health. In: Zenz C, ed. *Developments in occupational medicine*. Year Book Medical Publishers, Chicago, London 1980, pp 3–39.
5. Miettinen OS. Confounding and effect-modification. *Am j epidemiol* 100 (1974) 350–353.
6. Miettinen OS. Estimability and estimation in case-referent studies. *Am j epidemiol* 103 (1976) 226–235.
7. Miettinen OS. Design options in epidemiologic research – An update. *Scand j work environ health* 8 (1982): suppl 1, 7–14.
8. Miettinen OS. Causal and preventive interdependence. *Scand j work environ health* 8 (1982) 159–168.
9. Wang J-D, Miettinen OS. Occupational mortality studies: Principles of validity. *Scand j work environ health* 8 (1982) 153–158.

Appendix

The character of a cross-sectional case-referent study can be further elucidated through a consideration of the relationship between the incidence rate, IR , and the

prevalence rate, PR , usually taken as

$$PR = IR \times D \quad (\text{eq A})$$

where D is the duration of the disease.

Now it is required that the denominator of the incidence rate include also the person-years of the individuals suffering from the disease, ie, person-years of not being at risk. Distinguishing between the contribution of person-years from healthy and diseased individuals, equation A takes the form of

$$PR = [a/(Mt + aD)]D, \quad (\text{eq B})$$

where t denotes the experienced time of healthy individuals, M ; thus Mt and aD represent the total person-years of the base. Abbreviation with Mt results in

$$PR = (a/Mt)D/[1 + (a/Mt)D] \quad (\text{eq C})$$

$$PR = IR \times D/(1 + IR \times D), \quad (\text{eq D})$$

which may also be written as

$$IR \times D = PR/(1 - PR). \quad (\text{eq F})$$

Now the denominator of IR only encompasses person-years of healthy individuals, however.

Then, with the incidence rates among the exposed as IR_1 and among nonexposed

as IR_0 , the incidence rate ratio, according to equation F, may be written as

$$IR_1/IR_0 = [PR_1/(1 - PR_1)]/[PR_0/(1 - PR_0)] \quad (\text{eq G})$$

if it is presumed that the exposure does not influence the duration of the disease. Equation G may finally be rearranged as

$$IR_1/IR_0 = (PR_1/PR_0)/[(1 - PR_1)/(1 - PR_0)]. \quad (\text{eq H})$$

With the case-referent view, a and b may represent the exposed and nonexposed cases, ie, a/b represents (PR_1/PR_0) and, similarly, with c and d representing exposed and nonexposed referents (non-cases), c/d equals $(1 - PR_1)/(1 - PR_0)$ or

$$IR_1/IR_0 = (a/b)/(c/d) = ad/(bc). \quad (\text{eq I})$$

Thus the incidence rate ratio is obtainable from prevalent cases and noncases (referents) in the base population, given that the duration of the disease, D , is not influenced by the exposure. If the disease is rare, then the prevalence rate ratio comes close to the incidence rate ratio, as follows from equation H.

Received for publication: 4 January 1984