EPIDEMIOLOGY

EPIDEMIOLOGIC PRINCIPLES AND METHODS FOR OCCUPATIONAL HEALTH STUDIES

Carl M. Shy

A. DEFINITIONS AND USES

Definition and Scope of Epidemiology

Epidemiology is a study of the occurrence and distribution of disease in populations and of the factors that account for this distribution. Epidemiology shares with experimental and clinical medicine the overall objective of understanding causes of human disease. These three basic approaches to the study of human disease differ by the methodology each employs.

Experimental medicine, including the disciplines of microbiology, biochemistry, physiology, pharmacology, experimental pathology, and other basic medical sciences, utilizes the controlled experiment to test hypotheses about causal agents and disease mechanisms. It works with experimental models of human disease processes and brings to bear the powerful methods of controlled manipulation of variables and replication of results by different investigators. The greatest limitation of the experimental method is that it approaches the complex reality of human disease by isolating one variable after another within a framework of extremely simplified assumptions. This "scientific reductionism" often leads to conclusions that are removed from the overall causal chain of disease in man.

Clinical medicine and epidemiology, on the other hand, begin and end with disease in man, and both are more observational than experimental disciplines. Clinicians are concerned with individual diseased persons. They seek to diagnose the underlying disease that is causing the combination of symptoms, observable signs, and physiological and biochemical abnormalities detectable in affected individuals and to alleviate or mitigate the disease process or at least the pain and disability accompanying the disease. The clinician makes his diagnosis by gathering enough evidence about the patient to exclude all but one of the several disease entities that might account for the complex of clinical findings.

This reasoning process is largely based on empirical evidence reported in the medical literature and on an understanding of pathophysiologic mechanisms, rather than on general theories such as are available in the physical sciences. But the clinician's observations are based on a highly selected segment of the population, namely, those persons who seek medical attention. These persons are not necessarily representative of the population affected by occupational exposures.

Unlike the clinician, the epidemiologist does not usually have access to a wide array of clinical and biochemical information about a sick person. His immediate concern is not why an individual may be sick, but why disease frequency differs from one population to another, or from one time to another in the same population. The focus of epidemiology is with risk factors, often extrinsic to the sick person; it seeks to identify and quantify relationships between population groups at high risk and factors in the community or work environment that might account for the high risk.

Each of the three disciplinary approaches makes important and complementary contributions to our knowledge of human disease. The experimenter addresses disease mechanisms in an appropriate experimental mode, the clinician investigates disease manifestations in sick individuals, and the epidemiologist studies community determinants of disease risk. Disease treatment and prevention must be approached from each of these points of view, and the findings of one discipline can often lead to progress in the others.

Occupational Epidemiology

Occupational epidemiology is a study of the occupational environment as a risk factor for disease in working groups. The occupational setting is also used by the epidemiologist to obtain convenient access to populations in order to study coronary heart disease, bronchitis, high blood pressure, and other diseases that may not necessarily be primarily related to the work environment. The methods of occupational epidemiology are not generically different from those of acute or chronic disease epidemiology, but special features of the work environment are particularly beneficial to the epidemiologist. Some major advantages of occupational (group) epidemiologic studies are:

1. Complete plant populations can be readily constructed for previous years of employment.
2. Detailed individual exposure histories can sometimes be constructed from employment records.
3. In some plants, recurrent medical examinations provide sequential information on the health status of employees.
4. The vital status of an entire employment roster can be ascertained historically through retirement-insurance plans (though in many cases these plans cover only the vested worker).
5. In some plants, a single chemical dominates the exposure history of an occupational group, as in the case of vinyl chloride, nickel, chromates, asbestos.
6. Case-control studies conducted within a plant population can be referred back to a known population base, thereby allowing the investigator to obtain absolute estimates of risk and to evaluate the representativeness of the case and control study groups.

These features of occupational studies are important in the epidemiological approach to disease etiology. There are relatively few similar population settings in which the epidemiologist can as easily completely enumerate a cohort retrospectively, obtain detailed historical information on individual exposure, and simultaneously determine the vital status of the cohort. Much of our knowledge of chemical carcinogenicity in man has originated in studies of occupational cohorts.

A notable disadvantage of occupational studies is the fact that employed populations are usually healthier than the general population and, therefore, provide a biased representation of the true occurrence of disease in the entire community.

Uses of Epidemiology in Occupational Medicine

The occupational physician must depend on the skills of the clinician to detect disease occurrence in workers, but he needs the discipline of epidemiology to relate disease to factors in the occupational environment. To the extent that occupational medicine is concerned with prevention of hazardous occupational exposures, its basic science is that of epidemiology. There are, of course, other responsibilities requiring administrative and in some cases toxicological expertise, but insofar as the occupational physician wishes to approach his responsibility of disease prevention scientifically, he should be able to apply epidemiological methods in his practice. Unfortunately, the training of occupational physicians in the past failed to emphasize a rigorous curriculum in epidemiology and the related quantitative tools of the biostatistician.

Several uses of epidemiology have been described in the classical treatise by J. N. Morris (21). Among the uses most relevant to occupational medicine are the following, adapted from Morris' more generalized description:

1. To search for causes of disease and injury by comparing work exposures or other hazards of different occupational groups.
2. To study the history of disease patterns in occupational cohorts, describing changing patterns and possibly the changing character of disease, with a view to relating these changes to production and work processes.
3. To diagnose the health of the community of workers that fall under the purview of the occupational physician; to measure the magnitude and distribution of disease in terms of incidence, prevalence, disability, and mortality; to set occupational health problems in perspective

against other risk factors; to identify subgroups that require special surveillance and medical attention.

4. To evaluate the effectiveness of occupational health services with a view to an improved allocation of scarce medical resources, elimination of unnecessary practices, and introduction of new procedures that can be used to assess disease risks.
5. To identify new disease syndromes or disease entities related to the introduction of new agents or processes in the work environment, e.g., mesothelioma and asbestos exposure, angiosarcoma and vinyl chloride exposure.
6. To account for the entire spectrum of occupational disease risks, from the earliest preclinical manifestations in exposed workers to the development of latent disease excess by (a) including workers from first employment through those retired for many years; and (b) by following the course of disability and disease from first occurrence to subsequent etiology of a disease process.

As noted by Morris, these uses derive from the principle that epidemiology is a study of disease distributions and of the determinants of differences in these distributions in population groups. In systematically gathering information on disease distributions in occupational cohorts, the occupational physician has the opportunity to assess risk factors, evaluate occupational health services, describe changing patterns of disease in relation to work practices, and diagnose the health status of "his community." Epidemiology provides the practitioner of occupational medicine with the principles and methods to make valid assessments of possible associations between occupational exposure and disease risk.

B. EPIDEMIOLOGIC STRATEGIES, INDICES OF DISEASE AND MEASURES

General Notation

E = exposure or study factor, or persons exposed

$\overline{E}$ = absence of exposure, or persons not exposed

D = disease or death

$\overline{D}$ = absence of disease or death

PAR = population at risk

N = size of the study population

n = size of a subgroup in the study population

RF = a risk factor for disease, other than the study or exposure factor

CF = a confounding factor (to be defined subsequently)

EM = an effect modifier

I = incidence of disease

CI = cumulative incidence

ID = incidence density

Pr = prevalence of disease

r = rate of disease

P(D/E) = probability of disease, given exposure

RR = relative risk, the ratio of disease incidence in exposed to incidence in nonexposed

AR = attributable risk, or the difference in disease incidence in exposed and incidence in nonexposed

PrR = prevalence ratio, or the ratio of prevalence in exposed to prevalence in nonexposed

OR = odds ratio, an estimate of the relative risk

SMR = standardized mortality ratio (based on indirect adjustments for the distribution of other risk factors)

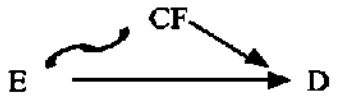
SRR = standardized mortality ratio (based on indirect adjustments for the distribution of other risk factors)

$\rightarrow$ = implies a causal association between a risk factor and disease

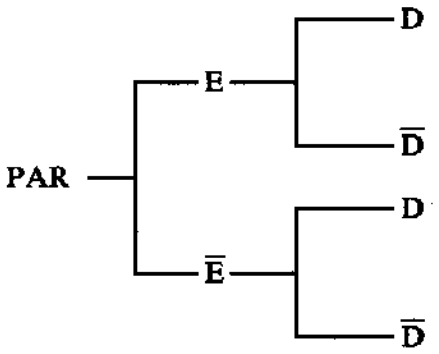
$\sim$ = implies a noncausal association in the distribution of two (risk) factors

Epidemiologic Strategies

The basic strategy of epidemiology is to establish an association (if one exists) between the distribution of group exposure to a study factor (E) and the distribution of disease (D), controlling for the presence of extraneous factors (CF = confounding factor) which may confound the relationship between D and E.



Symbolically, the epidemiologist works in the following framework:



Assuming that exposure and disease can be simply dichotomized, the framework of an epidemiologic study can be reduced to a 2 x 2 table:

	E	Ē	
D	a	b	m_1
D̄	c	d	m_0
	n_1	n_0	N

where m_1 is the number of persons diseased, m_0 the number without disease, n_1 the number exposed, n_0 the number unexposed, a the number of exposed persons with disease, b the number of unexposed persons with disease, etc.

If confounding factors are present, a separate 2 x 2 table must be constructed for each level of the confounder. For example, in the study of asbestos exposure (E) and lung cancer (D), a separate 2 x 2 table would be made for cigarette smokers and nonsmokers if smoking habits were unequally distributed between asbestos workers and others (and thereby created a situation of confounding). Confounding is controlled by stratification on the confounding factor, as will be discussed later.

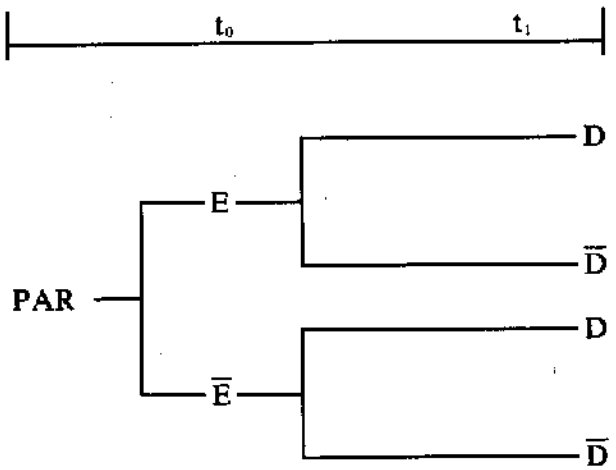
The initial step in an epidemiologic study of a work environment is to *describe* the distribution of disease (or functional impairment) among the working population of a plant or industry, without as yet postulating that a causal relationship exists between the work environment and disease. Thus, in the early studies of rubber workers (16) and steel workers (12), the investigators attempted to measure whether any specific disease excess could be found in the total cohort. These initial descriptive studies have their place in *generating hypothesis* for subsequent study, but they lack the necessary specificity and scientific rigor of hypothesis-testing investigations. Descriptive studies are of use in identifying high risk groups (e.g., cigarette smoking asbestos workers); in detecting temporal changes in disease frequency that might suggest causal agents; and in demonstrating whether there are geographical differences in disease distribution that might subsequently be explored for etiological significance. Descriptive studies are by nature epidemiologic "fishing expeditions" in which the first clues to population differences in disease distribution are obtained as warning signals that certain groups, places, or times deserve special attention. In some cases, clinical observations on disease clusters within a plant may raise the level of concern, but these observations need to be confirmed by some form of descriptive epidemiologic study in which disease frequency can be related to the working population at risk and compared with "expected" disease frequency.

The next phase of epidemiological investigation is the *analytical study,* designed with a specific testable hypothesis in mind. For example, is leukemia among rubber workers related to solvent exposure? Before an analytical study is initiated, the investigator must have a biologically plausible basis for postulating an association between E and D, and must usually have positive results from a descriptive epidemiologic study suggesting that a specific disease or cause of death is likely to be associated with a particular exposure. Analytical studies are definitive to the degree that they measure and

control for other known risk factors, but in the early stages of etiologic investigations it is often difficult to obtain detailed information about all the risk factors of interest. Thus, epidemiologic inferences are broadened usually by replication of results under different circumstances and by different investigators. This "consistency" characteristic of analytical epidemiology plays a key role in the extension of epidemiologic hypothesis to broader population groups, as will be discussed in the final section of this chapter.

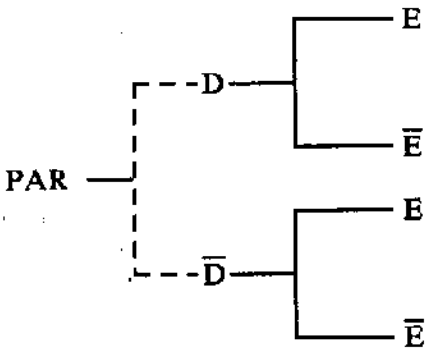
In addition to the descriptive or analytical nature of epidemiologic investigations, three basic epidemiologic study strategies can be identified. These strategies are distinguished by the temporal sequence in which exposure and disease characteristics are ascertained by the investigator.

1. The ***cohort study*** is a longitudinal progression in time from exposure to disease occurrence in populations at risk. Schematically, this approach is:



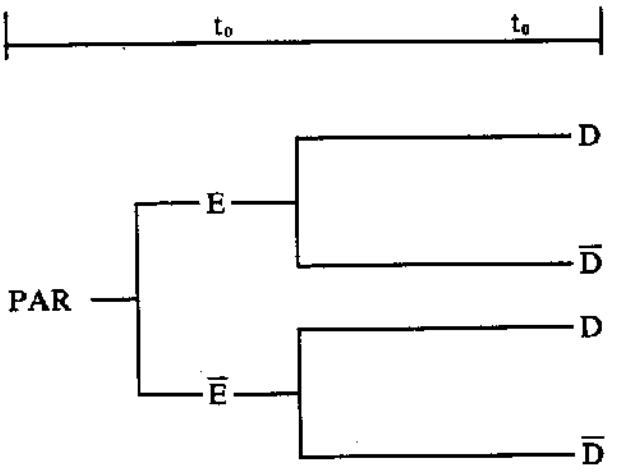
where t_0 is a time clearly preceding the occurrence of disease when exposure characteristics of the PAR are known, and t_1 is a subsequent time when new disease events have occurred in the E and $\bar{E}$ populations.

2. The ***case-control study*** is a retrospective progression from disease occurrence to exposure characteristics in D and $\bar{D}$ groups, usually without knowledge of the frequency of E or D in the source population at risk. Schematically:



The case-control study begins with the selection of cases and controls (without knowledge of absolute disease frequencies in the PAR). Subsequently, exposure and other risk factor information is sought for cases and controls.

3. The ***cross-sectional study*** consists of a simultaneous characterization of exposure and disease in a population at risk. Like the case-control study, this strategy is retrospective in approach. Schematically:



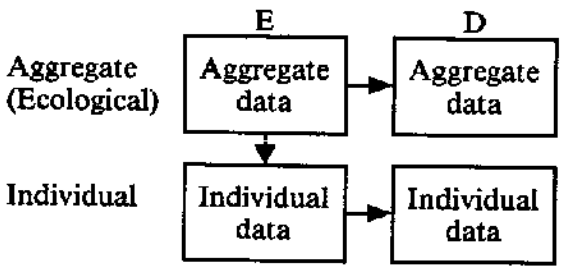
Unlike the cohort study, the cross-sectional investigation in and of itself pro-

vides no basis for ascertaining new disease events in E or $\bar{E}$ subgroups and is therefore unable to provide certain knowledge about antecedent-consequent relationships. A similar deficiency regarding temporal sequence applies to case-control studies. However, the cross sectional study does allow inferences about absolute levels of disease frequency in the source population.

The strengths and weaknesses of these three study strategies will be amplified in the following sections. It should be noted here that descriptive or analytical studies can be conducted within the context of any one of the three basic study strategies.

Epidemiologic investigations can also be distinguished by the nature of the linkage between exposure and disease. Schematically, there are two types of exposure-disease linkages possible:

Type of Exposure-Disease Linkage



In individual risk studies, knowledge is gained of the exposure characteristics of diseased and disease-free individuals. In aggregate studies (as exemplified by investigations of geographical variations in disease distributions), the researcher obtains data *separately* on the frequency of exposure and disease in that place. No data are available on the exposure or other risk factor characteristics of individuals who actually died or survived; thus, there is no evidence that death or disease occurred in exposed individuals. Readers will commonly assume that such a link exists, but the fallacy of this assumption—the "ecological fallacy"—lies in the fact that individuals who died of the disease may not have been actually exposed to the study factor, even though they lived in a place characterized by a high level of the exposure factor. For example, a small or large proportion of lung cancer deaths in counties having a petrochemical plant may be occurring in residents who had little or no occupational or environmental contact with the plant. Thus the aggregate study lacks linking evidence between exposure and disease at the level of an individual's experience and, to this extent, cannot establish causal relationships or quantify the magnitude of a risk factor for disease. The aggregate approach is intrinsically incapable of providing evidence useful for testing hypotheses about risk factors for disease and should be conceptually limited to the descriptive or hypothesis-generating category.

In summary, epidemiologic studies can be categorized on several different dimensions relating to the study hypothesis, the temporal sequence of exposure-disease ascertainment, and linkage between exposure and disease. (See diagram below.)

DIMENSION	TYPE OF STUDY
Elaboration of the hypothesis regarding E and D association	Descriptive (hypothesis generating) Analytical (hypothesis testing)
Temporal sequency of exposure —disease ascertainment	Cohort Case control Cross sectional
Nature of linkage between E and D	Individual Aggregate (ecological)

Epidemiological Indices

Indices of Disease Frequency

Two conceptually distinct measures of disease frequency are employed in epidemiologic studies: *proportions* and *rates*. A proportion is a ratio in which the numerator is a component of the denominator, e.g., the proportion of workers employed in 1970 that have retired by 1980, the prevalence of byssinosis in a textile plant, or the number of cases of lung cancer developing over a 10-year period in asbestos workers employed in 1965. In each case, the numerator is a count of persons who have or develop an event of interest such as retirement, disease, or death. The denominator contains the count of persons in the numerator plus all other persons who were in the same study group at the time the counting began. The value of a proportion can only range from 0 to 1, and because the units are the same (i.e., persons) in the numerator and denominator, they cancel out and the proportion becomes a dimensionless quantity.

The two most common proportions used to measure disease frequency in epidemiologic studies are *cumulative incidence* and *prevalence*. *Cumulative incidence* (CI) is a relatively recent term introduced to distinguish between the incidence measure that is a true proportion, i.e., cumulative incidence, from the incidence measure that is a rate, i.e., incidence density (see below). CI is a simple proportion of the study population that develops new disease events (new cases of disease, disability, or death). In the framework of the 2 x 2 table given earlier

	E	$\bar{E}$	
D	a	b	m_1
$\bar{D}$	c	d	m_0
	n_1	n_0	N

the CI in the exposed population is:

$$CI_E = a/n_1$$

and the CI in the unexposed population is

$$CI_{\bar{E}} = b/n_0.$$

In each case, n_1 and n_0 are a count of the number of exposed and unexposed persons at the beginning of the study. Thus, the cumulative incidence of lung cancer in a cohort of uranium workers, known to be alive and free of lung cancer in 1970, can be computed by following the cohort in time from 1970 to the termination of the study (e.g.,1980) and counting or accumulating all lung cancer cases developing in the cohort between 1970 and 1980. CI is meaningful only when the duration follow-up is given. *Prevalence* is also a proportional measurement, but it differs from cumulative incidence in that the numerator of a prevalence measure contains all diseased cases, whether new or old, that are "prevalent" at a point in time or during a time period. Thus a byssinosis prevalence of 0.20 signifies that 20% of the PAR was shown to have byssinosis, but no information is provided to determine whether the disease occurred recently or years ago. Prevalence is the measure of disease in cross-sectional studies.

Other epidemiologic indices that are proportions representing disease frequency are *case fatality rates* (proportion of cases that are fatal), *cumulative mortality* (proportion of a PAR that dies, usually computed for specific causes of death), and *proportional mortality* (proportion of all deaths due to a specific cause). In case-control studies, no direct measures of disease frequency can be computed, since these studies begin with the selection of cases and controls without direct reference to an underlying PAR.

The second distinct measure of disease frequency is the disease or death *rate*, which is defined as a measure of change in disease incidence per unit change in person-years at risk (during a specified time interval). This measure of disease frequency is termed the *incidence density* (ID), and its units are cases (or deaths) per person-years (or population-time). Like CI, ID is meaningful only when the time period is stated, e.g., per year. The term "incidence density" provides a specific name for the disease measure that allows exits and entrances to the study cohort, by virtue of deaths or losses to follow-up or by hirings of new workers during the course of the study. Thus, some persons in the population at risk will have been "at risk" during the entire duration of a cohort study, while others will have died early or entered late, so that their "at risk" experience is shorter than the former group. In the frame work of the 2 x 2 table, incidence density is given as follows:

	E	$\bar{E}$	
D	a	b	m_1
Person-yrs	N_1	N_0	

$ID_E = a/N_1$ $ID_{\bar{E}} = b/N_0$

N_1 and N_0 are not counts of exposed and unexposed persons but summations of the total time each member of the exposed and unexposed population remains in the study *and* free of disease, i.e., at risk.

A simple illustration will suffice to point out the difference between CI and ID. Assume that a cohort of 10,000 steelworkers is identified in 1970. Of these 10,000, 5,000 develop heart disease in the first year (an unrealistically high rate of disease). For ease of computation, assume that 1,250 cases occur on exactly each

terminal quarter of the year. Schematically, the PAR and deaths would be distributed as follows:

	t_0	$t_{1/4}$	$t_{1/2}$	$t_{3/4}$	t_1
PAR	10000	8750	7500	6250	5000
Cumulative cases	0	1250	2500	3750	5000

$$CI = \frac{\#\text{ cases by } t_1}{\#\text{ PAR at } t_0} = \frac{5000 \text{ persons}}{10000 \text{ persons}}$$
$$= 0.5 \text{ cases/person per year}$$

$$ID = \frac{\#\text{ cases by } t_1}{\#\text{ person-yrs. at risk}}$$
$$= \frac{5000}{10000(0.25) + 8750(0.25) + 7500(0.25) + 6250(0.25)}$$
$$= \frac{5000}{8125} = 0.61 \text{ cases/person-yr per year}$$

If all cases had occurred exactly on the first quarter date of the year the CI would be unchanged, but the ID would be:

$$ID = \frac{5000}{10000(.25) + 5000(.75)} = \frac{5000}{6250}$$
$$= 0.8 \text{ cases per person-yr per yr}$$

If 6000 cases had occurred on the first quarter of the year, the ID would be

$$\frac{6000}{10000(.25) + 4000(.75)} = \frac{6000}{5500} = 1.09$$

cases per person-yr. per yr. while the CI would be 0.6 per year. Thus ID expresses the average rate of case incidence in the true population at risk and takes into account not only how many cases occur but the "speed" at which these cases develop. ID can range in value from zero to infinity; it does not represent the probability of developing disease, as CI does, but rather the force of morbidity or mortality in a population. Since the ID is computed for persons only when they are actually enrolled in the study, the ID measure allows exits and entrances to the study population between the start and ending of the study.

Measures of Effect

If a causal relationship exists between exposure and disease, the measure of this relationship is the measure of effect. In cohort studies, measures of effect can be relative or absolute. The *relative risk* (RR) is the common relative measure and is given by the ratio

$$I_E / I_{\bar{E}}.$$

The absolute measure of effect is the *attributable risk* (AR)

$$I_E - I_{\bar{E}}.$$

Since there are two types of incidence measures, we can distinguish two types of relative and attributable risk measures as well: the *cumulative incidence ration* (CIR) and the *incidence density ration* (IDR) as measures of relative risk, and the *cumulative incidence difference* (CID) and the *incidence density difference* (IDD) as measures of attributable risk. These values can be computed from the typical 2 x 2 tables.

CI Study

	E	$\bar{E}$	
D	a	b	m_1
$\bar{D}$	c	d	m_0
	n_1	n_0	

$$CIR = \frac{CI_E}{CI_{\bar{E}}} = \frac{a/n_1}{b/n_0}$$

$$CID = a/n_1 - b/n_0$$

ID Study

	E	$\bar{E}$
D	a	b
person-yrs.	N_1	N_0

$$IDR = \frac{ID_E}{ID_{\bar{E}}} = \frac{a/N_1}{b/N_0}$$

$$IDD = a/N_1 - a/N_0$$

In the cumulative-type study, n_1 and n_0 represent the PAR at the study's inception. In the density-type study, N_1 and N_0 represent person-yrs. of follow up.

CIR and IDR (the relative measures of effect) are better indices of the strength of a potential causal relationship between E and D; CID and IDD are better indices of the impact on public health associated with exposure or the potential benefit of a prevention program in absolute numbers.

In cross-sectional studies, the prevalence ratio (PrR) is the only measure of effect commonly reported, and, like the CIR, is given by the ratio $\frac{a/n_1}{b/n_0}$. However, in this approach, a and b represent prevalent exposed and nonexposed cases representatively.

In case-control studies, the measure of effect is the odds ratio (OR). The odds ratio is an estimate of the relative risk of disease, given exposure. In a typical 2 x 2 table, the data layout for a case-control study is as follows:

	E	$\overline{E}$
D	a	b
$\overline{D}$	c	d
	a+c	b+d

In a cohort study, the measure of effect for these data would be:

$$RR = \frac{I_E}{I_E} = \frac{a/a+c}{b/b+d}$$

If the number of persons affected by disease is small relative to the number unaffected in the total population (the usual situation in studies of cause-specific diseases), then a + c is approximately equal to c, and b + d is approximately equal to d. Thus the estimate of relative risk

$$RR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

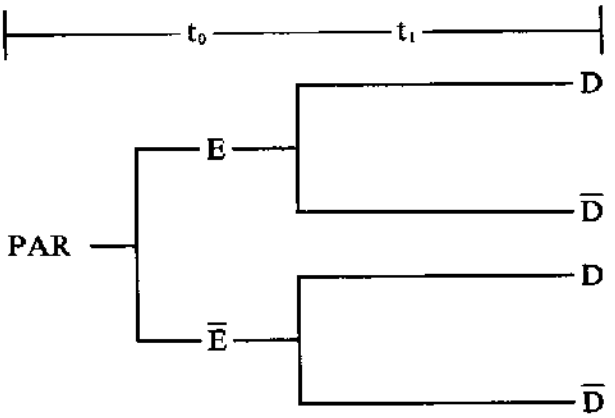
Thus $RR = OR = \frac{ad}{bc}$. The OR is a valid estimate of the relative risk derived from cohort studies provided two assumptions are met:

1. The disease is rare (thus a + c and b + d are reasonably approximated by c and d respectively in the general population).

2. Cases and controls are selected independently of exposure status or of any factor associated with exposure.

C. COHORT STUDIES

Characteristics of Cohort studies



The essential features of a cohort study are:

1. The study factor (E) is characterized in each person at t_0, prior to the appearance of disease.
2. The study population is observed (followed-up) longitudinally from t_0 to t_1; t_1 is determined by the onset time of disease or death, loss to follow-up, or cessation of the study.
3. New disease events occur between t_0 and t_1.
4. Measures of disease frequency can be referred to the PAR.

Cohort studies can be retrospective (or historical) or prospective, depending on the temporal relationship between the actual starting date of the study and the time when new disease events occur.

1. A retrospective cohort study: t_0 and t_1 have already occurred when the study is actually initiated by the investigator. This is the most common form of cohort study in occupational epidemiology.
2. A prospective cohort study: t_1 has not occurred when the study begins, and data on the cohort is first collected in real time, t_0. The study moves forward in real time, and new disease events are observed concurrently with the progress of the study.

Table I-44

LUNG CANCER IN COKE PLANT WORKERS BY LENGTH AND PLACE OF EMPLOYMENT

	Observed # Deaths	Observed/ Expected Deaths
Total coke plant	54	1.61*
5+ yrs in coke plant	46	2.09*
<5 yrs in coke plant	8	0.77
5+ yrs coke oven experience	40	3.67*
5+ yrs nonoven experience	5	0.51
5+ yrs oven topside	20	10.83*
5+ yrs oven side	18	2.49*

*$p < .01$

Source: Redmond et al. (23)

Examples

a. Retrospective cohort studies—Example: Long-term mortality study of steelworkers (12)(13)(23). In this series of studies, the mortality experience of nearly 60,000 steelworkers known to be alive and employed in 1953 in seven steel plants in Allegheny County, Pennsylvania, was followed retrospectively through 1966. Investigators determined the employment area of workers in 1953 from company employment records and compared the cause-specific observed mortality for a work area with expected mortality over the ensuing years, where "expected" was computed from the age and calendar year specific mortality of the total steelworker experience. Significant excess lung cancer mortality was reported for coke plant workers as shown in Table I-44. The observed/expected ratio was even greater among workers with five or more years employment in coke plants, with five or more years coke oven experience, and largest for workers with five or more years of topside oven work experience. This stratification of lung cancer deaths by work area revealed the fact that the small excess of lung cancer in the total coke plant was accounted for by men employed at the coke ovens, but that in this group, a tenfold lung cancer excess appeared in workers on coke oven tops where greatest exposure to coal carbonization by-products would be expected. The investigators computed "expected" mortality by applying age, race, cause-specific mortality rates in the entire steelworkers cohort to age, race, cause-specific mortality rates in the subgroup of the population at risk, in a given work area, in each calendar year of follow-up. This effectively derived an incidence density measure (i.e., # observed deaths/# expected deaths for the person-yrs. at risk in each calendar year)—a conventional calculation in retrospective cohort mortality studies. The expected incidence density was based on a comparison population consisting of a working population in the same industry and in the same geographical region as the exposed group. This approach avoided many of the selection bias problems encountered when national mortality data are used as a source for comparison with the mortality experience of a working population.

In a later report, Redmond and Breslin compared observed cause-specific mortality for the total Allegheny County steelworkers cohort against expected mortality derived from age, race, calendar year, and cause-specific U.S. mortality rates and Allegheny County mortality rates (Table I-45) (22). The authors noted that if U.S. rates are used as a basis of comparison, one would conclude that lung cancer is significantly in excess in both white and nonwhite steelworkers, whereas if Allegheny County rates are used as the baseline, lung cancer frequency is about the same in steelworkers as in the county's male population. It would, therefore, be erroneous to assume that the excess observed, when U.S. rates are applied, is directly related to occupational exposure. On the other hand, for many causes of deaths (such as cardiovascular and nonmalignant respiratory diseases) only overwhelming effects could be identified by using national or even regional data based on the general population's mortality experience. This phenomenon of apparent selection, at time of employment, of persons at lower risk for many causes of death has been termed the "healthy worker effect" and is a form of selection bias

Table I-45

OBSERVED/EXPECTED MORTALITY RATIOS BASED UPON U.S. AND ALLEGHENY COUNTY RATES FOR ALLEGHENY COUNTY STEELWORKERS

Cause of Death	Whites		Nonwhites	
	Based on U.S. Rates	Based on Allegheny Co. Rates	Based on U.S. Rates	Based on Allegheny Co. Rates
All causes	0.83*	0.77*	0.73	0.68
Cancer of lung	1.28*	1.04	1.64*	1.03
Cardiovascular and renal diseases	0.80*	0.74*	0.64*	0.64
Nonmalignant respiratory disease	0.61*	0.63*	0.72*	0.58*

*$p < .05$

Source: Redmond and Breslin (22)

that inadvertently occurs when the mortality experience of an occupational cohort is compared with that of the general population. The healthy worker effect appears to be greatest [cf. McMichael,(17)]:

1. at younger ages (less than 50-55 yrs.), when the selection process can better distinguish between the healthy and unhealthy.
2. for more overt disease manifestations such as cardiovascular and nonmalignant respiratory diseases, as opposed to cancer.
3. for nonwhites, whose employment opportunities often require better apparent physical health, or among whom a larger proportion of the general population may be in ill health relative to the employed.
4. for managerial, professional, business personnel, and department level supervisors than for clerical, semi-skilled operators, and unskilled workers.

b. Prospective cohort studies—Example: Lung cancer among uranium miners in the United States (1). Initial data for this investigation were provided from Public Health Service periodic medical surveys of uranium miners conducted during the period 1950 through 1960. Detailed occupational histories were obtained by personal interview at the time of each survey and were supplemented with annual uranium miner census information. Records were available for a study group of 3,366 white and 780 nonwhite miners who had one or more months of underground uranium mining experience prior to January 1, 1964. The mortality experience of this group was followed from date of first examination through September 30, 1968. The investigation thus possesses elements of a prospective cohort study, since employment and initial health data were obtained simultaneously with the initiation of the study by the Public Health Service. However the mortality experience of the cohort was largely assessed retrospectively, after the 1968 termination date for the ensuing mortality analysis. Information on vital status was obtained from records of the Social Security Administration (a common source for such information in occupational mortality studies), from the Veterans Administration, and through the annual census of miners, mail questionnaires, post offices, obituary notices, employment agencies, credit bureaus, and inquiry of local residents and relatives. As a result of this intensive follow-up program, the vital status of more than 99% of the cohort was determined. Likewise, the cumulative exposure of the miners to radon daughters was assessed retrospectively from 43,000 measurements made for approximately 2,500 uranium mines between 1951 and 1968. Cumulative radon daughter exposure values were calculated for each miner from the date of his first hiring to each sequential month of observation until termination of employment, death, or the 1968 cut-off date. The observed/expected mortality for lung cancer among miners is given in Table I-46, where "expected" is com-

puted from age, race, calendar year, cause-specific mortality rates for the male population of the four-state area in which miners were examined (Arizona, Colorado, New Mexico, Utah). In this case, the comparison population is a general, regional population, and the healthy worker effect will influence the interpretation. The effect of radiation exposure on lung cancer risk appeared to increase with higher cumulative doses, though the effect at lower doses is difficult to evaluate because as workers aged, their cumulative doses increased and their person-years of exposure were shifted to the next higher cumulative exposure category. It is, therefore, unknown whether workers in lower exposure categories would have experienced greater lung cancer mortality than expected had they left the industry and not accumulated further occupational radiation exposures. The cohort was further stratified on years after start of underground mining and into miners with and without previous experience in nonuranium hard rock mines. In each case, cumulative radiation exposure was shown to significantly increase the risk of lung cancer.

The relative advantages and disadvantages of retrospective and prospective cohort mortality studies are listed in Table I-47.

Criteria for Evaluating Cohort Studies

Cohort studies are conceptually straight forward approaches to assessment of disease risk in exposed workers. The incidence of disease can be directly compared in exposed and unexposed groups. However, as in any observational study there are a number of pitfalls that can invalidate the results of a cohort investigation. The following aspects of design and conduct of an occupational cohort study should be evaluated:

1. Were the criteria for an individual's entry into the study cohort completely described? A cohort is a population group possessing some common linking characteristic, such as being employed in the same plant on a certain date or in a specified time period. Criteria for entry to a cohort can be: age range, years of hire, membership in a union, employment status in a plant, etc. A loosely defined cohort will make it difficult to evaluated exposure status and consequent potential for disease in the total cohort. In accumulating person-years at risk, it is important not to mix persons of varying risk status into the same analysis pool.
2. What are the potential effects of nonresponse or refusal to participate in prospective cohort studies? If nonresponse is disproportionate among subgroups of exposed persons who are at a greater risk of disease (e.g., among asbestos workers who are cigarette smokers), the true risk of occupational exposure can be seriously underestimated. A well-designed study should provide some information, if only on a probability sample, about characteristics of nonresponders.
3. Is the exposure status of the "exposed" cohort uniform or heterogeneous? In most occupational environments, some workers are more exposed to the study factor than others in the same plant. How well could the investigators stratify the cohort on exposure potential? Pooling a heterogeneous exposure group will dilute the true risk of highly exposed with the low risk of relatively unexposed members of the cohort.

Table I-46

OBSERVED/EXPECTED LUNG CANCER DEATHS IN URANIUM MINERS ACCORDING TO CUMULATIVE RADON DAUGHTER DOSES

Estimated Cumulative Exposures (Working Level Months)	Person-yrs. at Risk	Observed/ Expected Lung Cancer Deaths
⩽120	8,516	0.55
120-359	9,355	4.67*
360-839	9,046	4.75*
840-1,799	6,607	4.76*
1,800-3,719	3,455	14.7*
3,720+	978	23.8*
Total	37,957	5.98*

*$p<.01$

Source: Archer et al. (1)

Table I-47

RELATIVE ADVANTAGES AND DISADVANTAGES OF THE TWO TYPES OF COHORT MORTALITY STUDIES

	Retrospective Cohort Mortality Studies	Prospective Cohort Mortality Studies
ADVANTAGES		
	1. Historical records are often available for complete enumeration of occupational cohorts.	1. Investigators can predetermine the kind of data they wish to obtain.
	2. Data are more readily accessible in a short time interval.	2. Data collection can be subjected to quality control.
	3. Lower cost.	3. Information on important covariables can be obtained.
	4. An efficient, feasible means to evaluate carcinogenic risks in industry.	4. Exposures can be directly measured, if necessary.
DISADVANTAGES		
	1. Information on important extraneous risk factors is often lacking.	1. High cost.
	2. Exposures must often be assesed indirectly, from employment records. Direct (instrumental) measurements of exposure are often lacking.	2. Time delay.
	3. Require relatively large sample sizes (thousands of person-yrs) for reasonable detection of disease risk.	3. Often infeasible due to time or cost constraints.
		4. Require relatively large sample sizes (thousands of person-yrs) for reasonable detection of disease risk.

4. How completely was the health or vital status of the cohort ascertained? Losses to follow-up greater than 10% subject a study to serious biases. A variety of standard techniques [cf. Boice (2)] are available to determine vital status, including searches of sources such as Social Security claims, vital registries of states, driver's license registrations, city and telephone directories, credit bureaus, contacts with former neighbors or fellow workers, etc. Until the national death index becomes operational, no one information source is adequate for follow-up of mortality status in the United States.
5. How valid is the selection of the "exposed" or comparison cohort? Several biases, such as the healthy worker effect, are possible in selecting a reference population. The disease risk of the ex-

Table I-48

LEAST SIGNIFICANT RELATIVE RISKS FOR VARIOUS SAMPLE SIZES IN A COHORT STUDY: TWO-SIDED SIGNIFICANT TESTS

Expected Annual Disease Rate in the Unexposed Group	Alpha Error	Person-yrs. of Follow-up per Exposure Group	Relative Risks at Beta Error of: 0.10	0.20
.01	.05	1,000	3.05	2.70
		10,000	1.51	1.44
		100,000	1.15	1.13
.01	.01	1,000	3.60	3.20
		10,000	1.62	1.54
		100,000	1.18	1.16
.001	.05	1,000	13.12	10.44
		10,000	3.07	2.70
		100,000	1.51	1.44
.0001	.05	1,000	>50	>50
		10,000	13.22	10.49
		100,000	3.07	2.71

Source: Walter (27)

posed and unexposed cohorts should be equal, except for the fact of exposure. To achieve this equality, the two cohorts must be stratified on extraneous risk factors for disease. This stratification may be impossible if the comparison population is inherently less or more healthy than those possessing the study factor. Several reference populations are available for occupational cohort studies. These include samples of the Social Security Administration files; other work groups, e.g., comparison of asbestos and nonasbestos textile workers (cf. Enterline (8)); and comparison of subgroups with the total occupational cohort, e.g., coke oven workers with all steel workers (cf. Lloyd (12)).

6. Was the size of the cohort large enough to detect a reasonable relative risk, i.e., what was the "power" of the study? Schlesselman (25) and Walter (27) provide tables and formulae for computing the sample size necessary to detect the smallest relative risk that can achieve statistical significance, given predetermined limits for alpha and beta errors and an expected frequency of disease in the unexposed population. As an example of the sample sizes required in cohort studies to detect various levels of significant relative risk, a portion of the calculations from Walter is reproduced in Table I-48 (27). The most important determinant of required sample sizes is the expected disease rate in the unexposed population. Studies of common diseases (such as cardiovascular disease) having an annual incidence rate of 0.01 could be designed with only 1,000 person-years of follow-up per group, to detect as significant a RR of 3.05. For cancer, specific sites in which the annual incidence might be 0.0001, 100,000 person-years of follow-up per group are required to detect a RR of 3.07. Note that these computations do not apply to study designs that utilize matching procedures and do not take into account stratification for various confounding factors.

Proportional Mortality Ratios

In some cases, all deaths occurring in a defined occupational cohort can be readily enumerated (e.g., through death claims against an employers' retirement system), but data are not as readily accessible on the size or composition of the population at risk. In these situations, neither cumulative incidence nor incidence density measures can be calculated. Instead, the relative frequencies of specific causes of death to total deaths (the proportional mortality ratio, or PMR) in the cohort can be compared with similar proportions computed for some comparison population such as the United States, the same state, or another occupational cohort. The PMRs can be adjusted for age differences in the 2 cohorts. Evidently, the sum of proportions for all causes will equal one in each group so that a relative excess for one cause in the study cohort will necessarily be offset by a deficit in other causes. The healthy worker effect and other problems affecting the validity of cohort studies will exist to the same degree in PMR studies. In addition, because of the offsetting problem already mentioned, it is likely that PMRs will suggest more deviations from the comparison population than will be detected by a true incidence study. Redmond and Breslin found 22 excesses or deficits in cause-specific proportional mortality of steelworkers by the PMR method, as opposed to 10 excesses or deficits detected by the standard cohort mortality study (22). The PMR method may be useful as a crude surveillance method, perhaps to suggest causes of death worth investigating in greater detail by the standard cohort or case-control study. However, the potential for false leads should be appreciated.

Standardized Mortality Ratio

The standardized mortality ratio (SMR) is the common summary measure of effect in occupational cohort mortality studies. This ratio is simply defined:

$$\text{SMR} = \frac{\text{Number of observed deaths in the exposed cohort}}{\text{Number of expected deaths in the exposed cohort}}$$

Where expected deaths are calculated by summing, overall ages, the product of the number of person-years for a specific age range in the study cohort and the cause-specific death rate in the same age range of the comparison population. Thus:

$$\text{SMR} = \frac{\Sigma \text{ observed deaths at age (i) in the exposed cohort}}{\Sigma \left(\begin{array}{c}\text{person-yrs in exposed} \\ \text{cohort at age (i)}\end{array}\right)\left(\begin{array}{c}\text{the death rate in the com-} \\ \text{parison cohort at age (i)}\end{array}\right)}$$

The purpose of the SMR calculation is to obtain a summary estimate of the mortality experience of the study cohort relative to the mortality experience of a comparison cohort of the same age composition. The SMR standardizes for age distributions or for any other risk factor that the investigator wishes to standardize on, such as calendar year, smoking habits if known, etc. There is, however, one serious limitation, frequently overlooked, in interpreting the absolute magnitude of an SMR. This limitation prevents one from comparing one SMR with any other SMR and thus from concluding that an SMR of 150, for example, indicates a greater mortality risk than an SMR of 125 in another cohort. This incomparability of SMRs can be illustrated with the hypothetical data presented in Table I-49, where two occupational cohorts, A and B, have different age distributions but identical age specific death rates. The SMR for A and B is based upon mortality rates in the same comparison population. Since the age specific death rates of A and B are identical, we expect the age adjusted summary value for mortality risk (the SMR) in the two cohorts to be equal. They are not. Close inspection of the formula for computing the denominator of the SMRs shows why the inequality occurred. The SMR value is weighted by the size of the age specific population in each study cohort. In cohort A, a large proportion of the population was older, and this age group experienced twice the mortality rate of the younger group. Thus a relatively high "expected" value was obtained for the denominator. In cohort B, the opposite distribution of the population by age yielded a relatively low expected value, thus a high SMR. SMR (A) differs from SMR (B) because we have used different weights—consisting of the age specific population size actually found in each cohort—in calculating the "standardized" mortality ratio. In effect, the adjustment for age is *internal* to the age structure of each cohort and is incomparable to a second SMR computed for a cohort with a different age structure. Since SMRs are computed to standardize on age struc-

Table I-49

NONCOMPARABILITY OF SMRs

	Cohort A			Cohort B			Comparison Population	
	No. person yrs.	Death rate	No. deaths	No. person yrs.	Death rate	No. deaths	Death rate	No. person yrs.
Age 40-49	1,000	.022	2	5,000	.002	10	.001	3,000
Age 50-59	5,000	.004	20	1,000	.004	4	.003	4,000
Total	6,000		22	6,000		14		

$$\text{SMR (A)} = \frac{22}{1{,}000(.001) + 5{,}000(.003)} = \frac{22}{16} = 1.375$$

$$\text{SMR (B)} = \frac{14}{5{,}000(.001) + 1{,}000(.003)} = \frac{14}{8} = 1.750$$

$$\text{SMR (A)} \neq \text{SMR (B)}$$

Source: Author: Hypothetical Data.

ture, it is apparent they are inefficient in this respect and effectively useless in cases where age structures of two populations are different.

Miettinen has proposed an external weighting scheme for risk factor adjustment (e.g., age) that avoids the flaw in comparing SMRs (20). This externally adjusted measure of effect is termed the standardized risk ratio (SRR), and is given by:

$$\text{SRR} = \frac{\Sigma \text{ expected deaths in the comparison population}}{\Sigma \text{ observed deaths in the comparison population}}$$

$$= \frac{\Sigma \left(\begin{matrix}\text{size of the comparison} \\ \text{population at age (i)}\end{matrix}\right)\left(\begin{matrix}\text{age specific death rate in} \\ \text{the study cohort at age (i)}\end{matrix}\right)}{\Sigma \left(\begin{matrix}\text{size of the comparison} \\ \text{population at age (i)}\end{matrix}\right)\left(\begin{matrix}\text{age specific death rate in the} \\ \text{comparison population at age (i)}\end{matrix}\right)}$$

Computation of the SRR, using the same data as given previously in Table I-49, is illustrated in Table I-50. The SRR (A) is identical to the SRR (B), and it should be. The identity is achieved by using as weights the "external" age distribution of the comparison population. The same set of weights is used in computing the SRR for cohorts A and B.

The bias of the healthy worker effect will operate in SRR as well as in SMR calculations, if the comparison population is a mixture of workers and non-workers. However, SRRs can be compared from one occupational cohort to another as long as they are based on the same comparison population. In some cases, a summary measure of risk should not be derived at all, particularly when inspection of age specific death rates reveals very different values among two or more occupational groups. In these cases, the SRR (or SMR) will average out or at least obliterate these age specific differences and mask the true nature of the risk differences between the cohorts. If a prominent difference in age specific death rates is found for two or more cohorts, these age specific rates should be reported, otherwise important etiologic clues may be entirely obscured.

D. CROSS-SECTIONAL STUDIES

The conceptual starting point of a cross-sectional study is a population or a representative sample of a population, such as all workers in a cotton textile plant. Typically, this population is divided into exposure groups, where exposure is characterized on the basis of current job assignments, current environmental monitoring, or other risk factors observed in the population at the time of the study. Exposure groups are simultaneously assessed for the presence or absence of disease, physiological abnormalities, or other health outcomes of interest that are prevalent in the population at the time of the study. For example, the presence of symptoms characteristic of byssinosis or of lung function abnormalities in textile workers, subdivided into cotton dust exposure categories, illustrates the

Table I-50

SRR: EXTERNALLY ADJUSTED MORTALITY RATIOS

$$\text{SRR} = \frac{\text{expected deaths in comparison population}}{\text{observed deaths in comparison population}}$$

$$\text{SRR (A)} = \frac{3000(.002) + 4000(.004)}{3000(.001) + 4000(.003)} = \frac{22}{15} = 1.467$$

$$\text{SRR (B)} = \frac{3000(.002) + 4000(.004)}{3000(.001) + 4000(.003)} = \frac{22}{15} = 1.467$$

$$\text{SRR (A)} = \text{SRR (B)}$$

Source: Author: Hypothetical Data.

cross-sectional design. Although the prevalence of disease at the time of study can be referred to a defined population at risk, as in cohort studies, the cross-sectional approach provides no data on new disease events (incidence data) or on the rate of disease development over time. To this extent cross-sectional studies are plagued by two inherent limitations concerning temporal relationship between exposure and disease.

1. The *antecedent-consequent relationship* of exposure and disease cannot be determined because exposed and nonexposed groups were not selected prior to development of disease.
2. The study population available to the investigators may be unrepresentative of the original exposed and nonexposed populations due to selective survival or selective migration of workers because of health reasons. Particularly in occupational settings, it is entirely possible that workers severely affected by their work environment may leave, be transferred to other jobs, or otherwise selectively drop out of the high exposure situation. This form of selection bias is illustrated with hypothetical data in Table I-51.

To counter these problems, it is possible to account for past job exposures and job changes of affected and unaffected workers. Also, observations may be made on early retirees, workers who transfer from hazardous work environments and others who leave a particular job category. However, information on job history is difficult to obtain by questionnaire techniques, and personnel records of the present employer usually provide no useful data on work histories from other plants. Further, because cross-sectional studies often involve large study populations, a complex work history file on each subject may be a costly data acquisition and data management problem.

The measure of disease frequency in a cross-sectional study is the prevalence (a proportion, not a rate) of affected persons in the population at risk. Prevalence is not a direct measure of disease risk in an exposed population because the nature of the study design does not generate incidence data. The prevalence of disease (or physiological abnormalities) in an exposed population is a function of two factors: the incidence and the duration of disease.

Prevalence = f (Incidence, Duration of Disease). A high prevalence may be brought about by a high incidence *or* by a long duration of disease. Cohorts enjoying better health care or favored treatment if illness develops may show a high prevalence of diseased workers, not because of high risk but due to longer "survival" of ill workers in the plant. Thus prevalence cannot be equated with incidence as a measure of disease frequency, and the prevalence ratio (PrR: ratio of disease prevalence in exposed to nonexposed) is not a wholly reliable estimate of risk associated with exposure. If however, incidence and duration are consistent over time or change equally in exposed and nonexposed groups, the PrR may be a valid indirect measure of relative risk (or cumulative incidence ratios). Unfortunately, there is seldom a basis for making these assumptions concerning change over time.

In many cases the nature of the occupational health problem is such that cross-sectional studies and generation of prevalence data are the only practical options available to an investigator. Such would be the situation where cumulative occupational exposures lead to increased risk of developing a chronic disease of insidious onset, such as chronic bronchitis or byssinosis. It is difficult to determine when these chronic respiratory diseases really begin and, for a retrospective cohort, to be certain who was free of the disease at some predetermined point in the past. Likewise, the cumulative exposure of working subgroups is usually difficult to evaluate. Ideally, one would like to begin with a standardized health examination that ascertains the presence or absence of chronic respiratory disease at the time of employment; follow different exposure groups serially with repeat health examinations; and finally assess the health status

of each exposure group at the study's termination. Unfortunately, such data are rarely generated, and selective losses of ill persons may still occur, although it would be possible to retrospectively evaluate the health status of the drop-outs and compare this with survivors in the same group.

Cross-sectional data are sometimes used to make geographical comparisons of disease prevalence between different countries, states, counties, or cities. These findings provide an index of the relative magnitude of a problem in different geographical areas and may be important in assessing the need for health care facilities and other resources. The prevalence of physiological abnormalities, such as impaired lung function or high blood leads, may provide the first clues to the existence of a work hazard. However, prevalence data should not be used to estimate disease risk unless there is reason to believe incidence and duration are relatively constant in exposed and nonexposed groups. Special efforts should be made to evaluate the possibility of selective survival or migration before drawing conclusions based on prevalence ratios.

Mortality Rates

Mortality rates have features of incidence and prevalence data in that the mortality rate in a given year is determined by the incidence, the duration or chronicity, and the virulence of the disease. For diseases such as lung cancer that are highly fatal in a relatively short time, mortality rates are reasonable indices of incidence rates. For avirulent diseases such as skin cancer, chronic musculoskeletal disorders, etc., mortality rates are totally unrepresentative of disease incidence. If persons survive for long periods with the disease—such as is the case with chronic respiratory or cardiovascular disease—mortality rates again do not reflect incidence unless survival (duration of disease) is relatively constant. Survival may be affected by temporal changes in medical care, by age at onset of disease, and by competing risks of death. Comparison of mortality rates in different geographical or occupational groups, when the mortality data were obtained outside the framework of a true cohort study, shares many of the limitations of prevalence data and should be interpreted similarly.

E. CASE-CONTROL STUDIES

Unlike cohort and cross-sectional studies, case-control studies are not inherently population based, do not conceptually begin with exposed and unexposed groups, and provide no direct measures of disease frequency in a population. In spite of these negative aspects, the case-control study has become an efficient, powerful epidemiological tool that is finding increasing applications in etiologic research, particularly in the area of occupational and environmental cancer research. The case-control study may not seem to be a direct and logical approach to evaluating the risk of disease in exposed persons; however Cornfield (7) and Mantel and Haenszel (15) provided a theoretical basis for estimating relative risks from case-control data and the strength of this method is being increasingly appreciated.

A case-control study begins with the identification of all disease cases that can be found in a source population such as a hospital register, a state tumor registry, the insurance files of an employer, or in a population-based disease survey. Controls, defined as persons not known to have the disease, are selected from the same source population and are often matched to cases on basic demographic factors such as age, sex, and race. For efficiency, the size of the control group is usually equal to, or a small multiple of, the size of the case group. As a result, case-control studies typically have sample sizes of 200 to less than 1,000 subjects in total, whereas cohort and cross sectional study designs frequently involve many thousands of individuals or person years.

Analytical Aspects

The scheme for a case-control study is as follows:

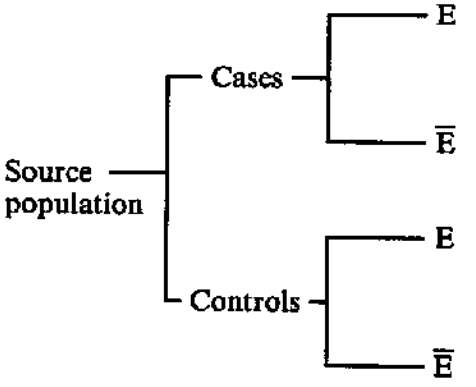
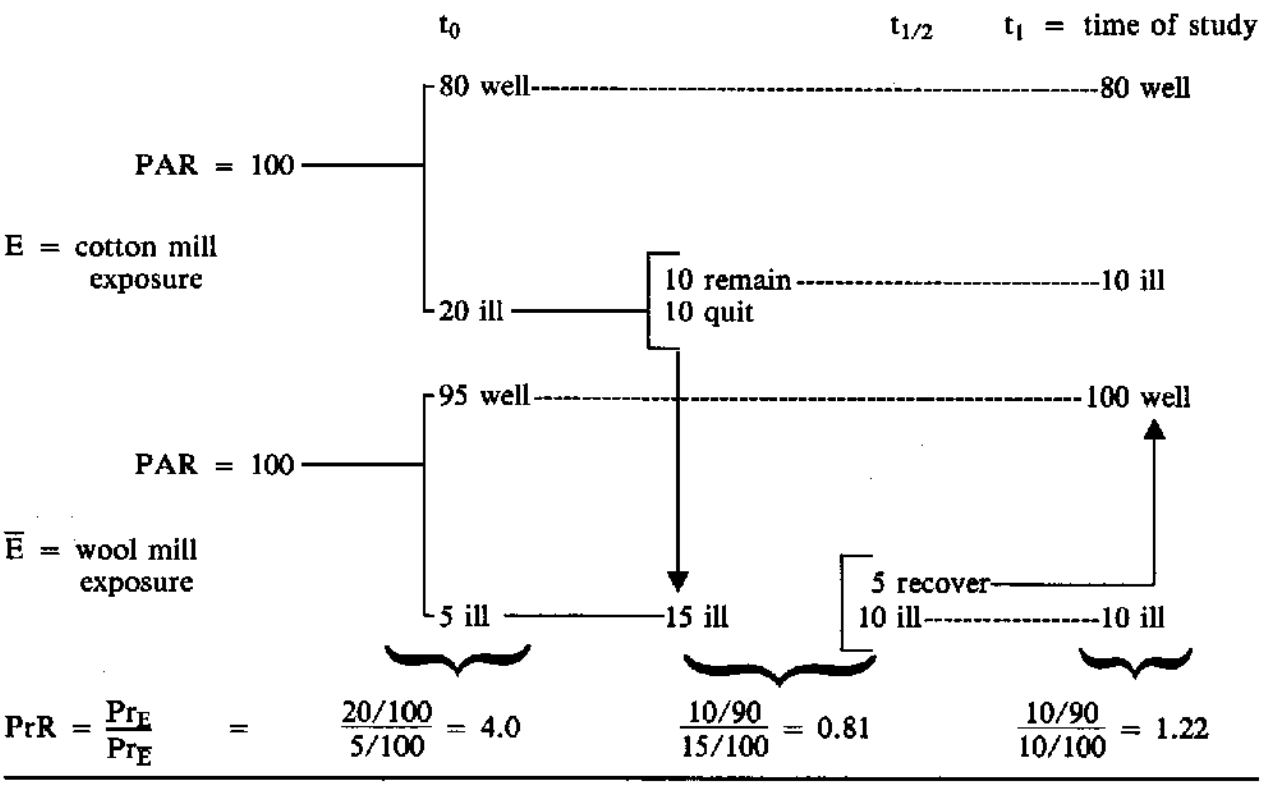


Table I-51

EFFECT OF SELECTIVE MIGRATION ON PREVALENCE RATIOS FOR RESPIRATORY DISEASE IN TEXTILE WORKERS



The initial (t_0) prevalence ratio (PrR) shows a fourfold greater prevalence of respiratory disease in cotton versus wool mill workers. Selective losses and migration of affected workers from cotton mills to wool mills results in a PrR of 0.81 at time $t_{1/2}$ (before the study actually begins). Recovery of ill persons due to cessation of exposure results in the PrR of 1.22 at t_1, (the time when the cross-sectional survey is actually conducted).
Source: Author: Hypothetical data.

	D	$\bar{D}$	
E	a	b	N_1
$\bar{E}$	c	d	N_2
	M_1	M_2	T

Exposure and other risk factor statuses of cases and controls are ascertained retrospectively. It is crucial to the study's validity that cases and controls be selected independently of exposure status. This independent selection can be a serious problem when cases are drawn from a source that is inherently at higher risk of disease (e.g., a hospital register) than the source for controls (e.g., the general population). Unlike cohort studies in which N_1 and N_2 (in the above 2 x 2 table) are fixed (not random) at the start of the study, the number of cases, M_1, and the number of controls, M_2, are fixed while the outcome of interest is the exposure distribution among cases (a/M_1) and among controls (b/M_2). Having determined the cell frequencies a, b, c, d, in the 2×2 table, the odds ratio (OR) can be simply computed, as discussed in Section B, and is given by ad/bc. As demonstrated by Cornfield, the OR is a valid estimate of the relative risk of disease, given exposure (7). Mantel and Haenszel provide methods to compute the statistical significance of an OR or of an RR (15). For a case-control study, the significance of OR can be computed by applying the Mantel-

Haenszel X^2 test with one degree of freedom, where:

$$X^2_{MH(1)} = \frac{(N_1 + N_2 - 1)(ad - bc)^2}{N_1 N_2 M_1 M_2}$$

In case-control studies where each case and control is individually matched on a factor such as age, sex, and race, the 2 x 2 table takes a different form from that of nonpaired-matched studies. The exposure status of each case-control pair is considered and entered into the appropriate cell of a matched pair 2 x 2 table as follows:

Matched Pair 2 x 2 Table
Case-Control Study

		Controls: E	Controls: $\bar{E}$
Cases: E		r	s
Cases: $\bar{E}$		t	u

In this table, r is a count of the pairs in which both the case and control are exposed, s is a count of the pairs in which cases are exposed and controls are unexposed, etc. In the matched pairs analysis,

$$OR = \frac{s}{t}$$

$$X^2 = \frac{(|t - s| - 1)^2}{t + s} \text{ (McNemar's Test)}$$

Design Aspects

In selecting cases, it is a distinct advantage to limit the case population to recently diagnosed or incident cases. Incident cases provide a more clearer differentiation between factors that influence disease etiology as opposed to those related to the duration and course of disease. Older cases still surviving are less representative of the population of cases in the source population. Incident cases offer greater potential for direct interviews or other means of acquiring fresh data concerning past exposures and other risk factor information.

It is desirable, though not essential to the internal validity of a study, that cases and controls be representative of the source population at large from which cases were derived. For cases, the best way to assure representativeness is to include all cases that are known to have occurred in the source population within a defined time period. The following sources of cases have been utilized:

1. Hospital registers. Cole et al. enumerated all cases of bladder cancer reported in eastern Massachusetts hospitals during the 18 months ending June 30, 1968; obtained a probability sample of matched controls from the general population; and related the findings to employment in various industries (5).
2. State vital statistics registers. Brinton et al. identified, from state vital records, all cancers of the nasal cavity and sinuses occurring between 1956 and 1974 in North Carolina counties in which at least 1% of the population was employed in furniture and fixtures manufacture according to the 1963 U.S. Census of Manufacturers (3).
3. Occupational cohorts. McMichael et al. evaluated job titles of all cancer cases that were identified during the course of a retrospective cohort mortality study of four rubber plants (16). Cases occurred between 1964 and 1973.
4. Tumor registries. The National Cancer Institute has initiated a large case-control study of bladder cancer and saccharin use. Cases are being obtained from the 10 United States cancer registries that form part of the SEER Program (Surveillance, Epidemiology and End Results).

Each of the above sources of cases provides distinct advantages and disadvantages. Hospital registries generally are more accessible to researchers and allow easy validation of case reports against biopsy or autopsy evidence. However, a single hospital may not draw its patients from a clearly defined source population, and some cases of the disease from this population may go elsewhere, or not seek medical attention at all. Incident cases can most easily be identified through hospital registers. Vital records assure a nearly complete enumeration of cases, providing the disease is listed on the death cer-

tificate and sufficient time is allowed for nearly all diseased persons to have died. Hence, incident cases cannot be obtained from vital registries. Inaccuracies of diagnosis on death certificates must be assessed by linking death records to hospital records, a time consuming and logistically difficult procedure. Tumor registries, especially if state wide, offer nearly complete enumeration of incident cases, with reasonably good confirmation of diagnosis based on tissue samples. Unfortunately, there are very few comprehensive state-wide tumor registries in the United States. The hybrid design of a case-control study nested within an occupational cohort study is a recent method applied in studies of U.S. rubber workers [cf. McMichael (16)]. This approach affords a clear frame of reference to the source population from which absolute measures of disease risk can be derived. The design allows the investigators to assess how well controls are representative of the source population.

Selection of controls for case-control studies is a difficult epidemiologic issue. MacMahon cites several concerns in selection of controls (14):

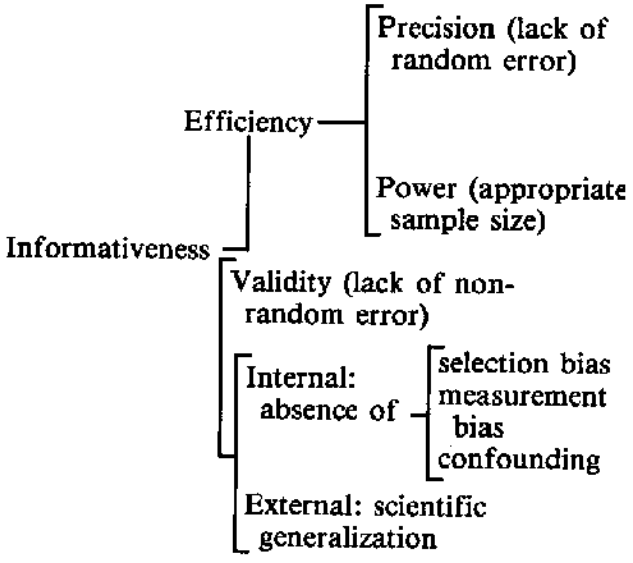
1. Controls should be representative of the source population at large from which cases were derived. The surest approach is to draw a probability sample of all noncases in the source population, but this is rarely feasible. Cole et al. obtained controls for their eastern Massachusetts bladder cancer study by having access to a published listing of all adult residents stratified by age and sex (5). Probability samples of the dead population of a state can also be obtained from vital registries. Controls drawn from the hybrid case-control-within-cohort design can also be obtained by probability sampling.
2. Information on exposure and other risk factors should be obtained with the same degree of accuracy and ease for both cases and controls. The problem is that cases may be so concerned that they (or their relatives) exhibit selective recall of past exposure or risk factor experiences. Live controls may provide better information on their own personal habits and employment histories than relatives of dead cases.
3. Controls should be similar to cases with respect to generally recognized, potentially confounding factors. Controls drawn from a different source population, e.g., hospital cases versus community controls, may differ in the distribution of other risk factors. These differences can be controlled by matching cases and controls in the selection process and a subsequent matched-pairs analysis or, after selection, by stratification analyses. Though there is considerable discussion in the literature on the advantages and disadvantages of matching in the selection process, most investigators agree that some form of stratification analysis is necessary to control for confounding [cf. Mantel and Haenszel (15)]. Matching during selection places constraints on what controls can be included and is perhaps most justifiable when it is very costly to obtain exposure and other risk factor information from cases and controls. Individual matching followed by matched-pairs analysis assures that cases and controls will be similar with respect to potential confounders that are the basis for matching. Thus, matching in the selection process assures that all cases and controls will provide useful information, while matching by stratification after analysis may cause some losses of unmatchable cases or controls. Since it is seldom possible to match on all important potential confounders, matched pairs must often be disaggregated in order to perform a stratified analysis controlling for several confounders simultaneously. Over-matching in the design phase occurs when cases and controls are matched on variables that are not risk factors for disease or when subjects are matched on variables that are intermediate in the causal pathway (e.g., matching on lung function would minimize the likelihood of detecting a smoking effect). Similarly, if cases and controls are matched on county or city boundaries, their general environment (air and water quality) may be so similar that effects of certain aspects of environmental quality could not be detected.

In summary, the major advantages of case-control studies are: efficiency in terms of relatively small sample sizes required to detect minimum risks, ability to access and process more detailed information on individual exposure and other risk factors of interest, and a reasonable time frame for completion of studies.

The disadvantages are: risk of selection bias, difficulty of obtaining controls representative of the source population, difficulty of getting equally reliable information from cases and controls. The hybrid case-control within a cohort design is a promising method that overcomes some of these disadvantages and is particularly applicable for occupational studies of certain disease risks.

F. SOURCES OF ERROR IN EPIDEMIOLOGICAL STUDIES

Two major types of error in observational studies—random (sampling) and nonrandom (systematic) error—cause loss of information in epidemiologic data due respectively to loss of precision (efficiency) or loss of validity. The informativeness of a study may be considered as follows:



An informative epidemiologic study is one which can efficiently detect an association between exposure and disease, if the association truly exists, and which can provide a valid estimate of the association's magnitude.

Efficiency

Efficiency refers to concerns about random sampling, sample sizes, and statistical inferences from results obtained in the study population to conclusions about exposure-disease associations in the source population. Statistical *precision*, one component of efficiency, is a measure of the variability of repeated measurements of the same phenomenon, e.g., the incidence of lung cancer in the asbestos industry, or the prevalence of byssinosis in the textile industry. A better estimate of disease frequency will be obtained if some form of random sampling is used, and if the sample size is large enough to represent the source population, e.g., the plant or the industry. Thus, the precision of a study can be enhanced by increasing the sample size and obtaining better probability samples of the source population. Precision is inversely proportional to the variance of the estimate and thus to the confidence interval about the point estimate of the measure of effect, i.e., the relative risk, prevalence ratio, or odds ratio.

The power of a study refers to the adequacy of the sample size for detecting an effect, if one exists, at a certain minimum relative risk. If a tenfold disease excess in lung cancer exists among asbestos workers, a considerably smaller cohort can be studied than if a twofold excess were expected. A negative study may be accepted as an adequate assessment of exposure-disease relationships only if the sample size was large enough to detect a predetermined level of effect such as a twofold relative risk. If the sample size was sufficient only to detect a fourfold or fivefold relative risk, then a negative result may have little meaning.

Validity

Assuming that random or sampling error is reasonably controlled, a study can still yield an erroneous conclusion concerning the existence and magnitude of an association between exposure and disease in the source population. That is, the estimate of effect can be distorted by several systematic or nonrandom errors, and these errors are usually termed "biases." An example is the bias of the healthy worker effect, or of comparing hospital cases with community controls, or of confounding due to the mixture of exposure with another risk factor for disease. These errors are systematic in that they cause a

unidirectional deviation of the measure of effect toward or away from the null hypothesis of no effect.

The internal validity of a study refers to the agreement between an estimate of effect derived from a study sample and the level of effect that actually exists in the source population. Internal validity is distinct from statistical precision, and the distinction can be illustrated by an analogy. If 20 darts are thrown at a bull's eye, the spread and accuracy of the darts around the bull's eye may be characterized by one of four combinations:

		Spread = Precision (lack of random error)	
		P	$\overline{P}$
On Target = Validity (lack of nonrandom error)	V	VP	$V\overline{P}$
	$\overline{V}$	$\overline{V}P$	$\overline{V}\,\overline{P}$

VP = all darts close together and on target (precise and valid)

$V\overline{P}$ = darts are spread but center on the target (imprecise and valid)

$\overline{V}P$ = darts are close together but off target (precise and invalid)

$\overline{V}\,\overline{P}$ = darts are spread and center off target (imprecise and invalid)

Thus an odds ratio of 3.0 could be a valid estimate of the risk of leukemia associated with solvent exposure in dry cleaning plants, but, due to imprecision, the confidence interval about this point estimate could be so large as not to be statistically significant. On the other hand, an OR of 3.0 may be statistically significant but invalid due to confounding of solvent exposure with another risk factor for leukemia.

A second form of validity, external validity, refers to the ability to generalize as, for example, from the study of solvent exposure in a dry cleaning plant to solvent exposure under other circumstances (in other industries, other solvents, etc.). External validity is evaluated by a complexity of criteria that include considerations of consistency with other studies, biological plausibility, convergence of evidence from several biological disciplines, knowledge of pathophysiological mechanisms, evidence from experimental investigations, etc. Hence, while a study may be internally valid, the ability to make scientific generalizations may be sharply limited by our lack of knowledge about the biological mechanism of the effect or the circumstances that may modify the association between exposure and effect. A single epidemiological study cannot be definitive on most of these issues, largely because of the complex nature of human responses to the total environment. External validity is a function of the breadth and depth of knowledge brought to bear on a subject by all biological disciplines, and it is within this context that questions of causality must be addressed.

Three sources of nonrandom error need to be considered in depth, since they are pervasive sources of bias in nearly all epidemiologic studies.

Selection Bias

Selection bias is a distortion in the estimated measure of effect due to the influence of the outcome variable (i.e., disease frequency in cohort studies, exposure frequency in case-control studies) on the selection of subjects into the study. For example, air pollution induced diseases may cause ill persons living in polluted areas to migrate to less polluted communities. Several forms of selection bias have been illustrated in previous sections:

1. The healthy worker effect: a selection bias that operates when mortality or morbidity of a working group is compared with that of the general population, components of which have poor health status.
2. Selective migration or survival: differential movement of persons affected by their exposures to less hazardous environments, such movement taking place prior to initiation of a study; survival of the healthier segment of a population exposed to an environmental hazard.
3. Selective losses to follow-up: disproportionate losses from a cohort of persons who are exposed and become ill.
4. The short-term worker effect: a phenomenon whereby short-term workers who move from one employer to another are often found to have below average health status and above average mortality. Failure to account for these workers biases

the estimated association between level of exposure (as indexed by duration of employment) and disease risk. Industries with high labor turnover are particularly subject to this bias.

5. Case-control biases. A number of subtle selection biases can operate to cause exposed cases to be more readily included in a study than exposed noncases. As an example, if cases of breast cancer were obtained from a screening clinic and controls from a community, a case-control study of birth control pills as a risk factor may be biased by the fact that pill users are more carefully watched for complications and therefore sent to breast cancer clinics. Similarly, hospitalized patients are, in general, more likely to be smokers and users of medications than community controls. The potential for selecting exposure-disease combinations needs to be carefully assessed when the results of any case-control study are interpreted. For a recent debate on this issue, refer to Horwitz and Feinstein (10) and a rebuttal by Hutchinson and Rothman (11). At times, an empirical approach may be taken to avoid selection bias in case-control studies, whereby dual controls are selected: one from the general source population (a "loose" control) and one from a population that is more closely matched to cases on potential confounders such as use of health care facilities, or date of hire (a "tight" control). Note: a more complete discussion of potential biases in case-control studies is given by Sackett (24).

   Selection bias is more likely to be a problem in case-control and prevalence than in cohort studies. Cohort studies by definition begin with disease-free individuals, whether exposed or not. Of course, disease-prone individuals could have selected themselves out of the exposure cohort prior to the initiation of the study, but generally disease risk is not perceived differentially between exposed and unexposed groups.

To cope with selection bias, investigators can take several measures in the design and analysis phase of the study:

1. In the study design:
   a. Reduce losses to follow-up in cohort studies by intensive follow-up efforts.
   b. Reduce nonresponse rates or obtain information on a sample of nonrespondents.
   c. Carefully select controls for cohort and case-control studies to assure that, under the null hypothesis of no effect, controls have the same risk as cases or that exposure status does not differentially influence the selection of cases and controls.
   d. Make special efforts to obtain historical information on a sample of persons who departed from a plant or geographical area prior to the initiation of the study.
2. In analysis:
   a. Try to estimate the direction of selection bias by analyzing data on a sample of nonrespondents or on "reluctant" versus "willing" responders.
   b. Compare whatever is known about those lost, versus not lost, to follow-up.
   c. Estimate the extreme situation for effect of losses to follow-up, namely that all losses from the exposed group remain disease-free, while losses from the nonexposed develop the disease.

Measurement Bias

Measurement bias is a distortion in the estimated measure of effect, due to errors in measuring exposure or disease status or to misclassification of subjects with respect to exposure or disease status.

Sources of measurement error include:

1. Variation among observers or instruments, or internal variation within the same observer or instrument: e.g., well trained radiologists may differently interpret the same chest roentgenogram.
2. Variation in the subject or exposure situation being measured, where our limited measurement systems fail to adequately represent these variations; e.g., one blood pressure reading is taken to represent an individual's blood pressure

status even though he may exhibit diurnal variations.

No instrument or observer can obtain perfect measurements at all times. Measurement bias refers to systematic, rather than random, errors associated with the taking of measurements. The epidemiologist uses two different but related terms to assess the presence of systematic error in measurements: *sensitivity* and *specificity*. Sensitivity is the proportion of true cases (or true exposures) detected as cases or exposed by a test, an observer, or an instrument. Specificity is the proportion of true noncases (or nonexposures) detected as noncases or nonexposed by a test, an observer, or an instrument. The concepts are well illustrated in a 2 x 2 table:

Test for presence of disease (or exposure)		Actual Disease (or) Exposure) D	$\overline{D}$
	D	a	b
	$\overline{D}$	c	d
		a + c	b + d

Se = sensitivity = a/a + b
Sp = specificity = d/b + d

Related to these measures are:
FN = % false negatives = c/a + b
= 1 − sensitivity
FP = % false positives = b/b + d
= 1 − specificity

Note that sensitivity provides information about persons with disease (or exposure), whereas specificity applies to persons free of disease (or to the nonexposed). In order to obtain estimates of sensitivity and specificity, it is necessary to obtain measurements of the same event (disease or exposure) by means of the usual test or instrument and by a second, more complete or accurate method that would be considered the standard of excellence. For example, one could test for chronic respiratory disease with ventilatory function and/or a form of the standardized chronic respiratory disease questionnaire. The same persons could then be carefully examined by a panel of experts who might perform a battery of diagnostic procedures, pool their findings, and attempt to reach diagnostic agreement. Similarly, an area monitor in a work place might be compared with results from all individuals wearing personal monitors, combined with careful industrial hygiene evaluations, to assign an exposure value to a given job.

Although sensitivity and specificity measures are seldom obtained for most diagnostic and screening tests or for environmental monitors, users of these test instruments have the opportunity to develop their own validation procedures. The importance of sensitivity and specificity measures lies in the application of these measures to the assessment of measurement bias and the potential for obtaining corrected estimates of effect, once sensitivity and specificity are known. This use of sensitivity and specificity has not received the attention it deserves, and the importance of the point will be illustrated in the following:

1. Estimating the magnitude of information bias:

Assume that the following results are obtained from a cohort study:

	D	$\overline{D}$	
E	a = 100	b = 400	$N_1 = 500$
$\overline{E}$	c = 50	d = 450	$N_2 = 500$
	$M_1 = 150$	$M_2 = 850$	T = 1000

Cumulative Incidence Ratio (CIR)

$$= \frac{a/N_1}{c/N_2} = \frac{100/500}{50/500} = 2.0$$

Assume that the method for measuring exposure status can be shown to have a sensitivity of 90%, and a specificity of 90%, and that this measurement error is equal for diseased and nondiseased groups. Applying a 90% sensitivity (Se) and specificity (Sp) to the diseased and nondiseased groups separately, we obtain the following:

Diseased

Actual exposure status

	E	$\overline{E}$	
E	a 'Se	c '(1 − Sp)	100 = a
$\overline{E}$	a '(1 − Se)	c 'Sp	50 = c
	a '	c '	
	Se = 0.9	Sp = 0.9	$M_1 = 150$

Nondiseased

Actual exposure status

	E	$\bar{E}$	
E	b 'Se	d '(1 − Sp)	400 = b
$\bar{E}$	b '(1 − Se)	d 'Sp	400 = d
	b '	d '	
	Se = 0.9	Sp = 0.9	M_2 = 850

The values a ', b ', c ', and d ', which are baseline marginals for the two 2 x 2 tables representing diseased and nondiseased subjects respectively, are the true exposure frequencies:

a ' = actual number of exposed diseased subjects

b ' = actual number of exposed nondiseased subjects

These values can be calculated, first by applying the known Se and Sp measures to the unknowns, a ', b ', c ', and d ', yielding the value given in the cells of the 2 x 2 tables immediately above.

It can be shown [cf.Shy et al. (26) and Copeland et al. (6)] that it is possible to solve for a ', b ', c ', and d ' in terms of Se, Sp, M_1, M_2, N_1, N_2, T, a, b, c, and d. Knowing the values of a ', b ', c ', and d ', we can calculate the true cumulative incidence ratio as follows:

For misclassification of exposure status

$$\text{True CIR} = \frac{a'/N_1}{c'/N_2} = \left(\frac{M_1\,Sp - c}{M_1\,Se - a}\right)\left(\frac{TSe - N_1}{TSp - N_2}\right)$$

Note: this formula applies to errors in measurement of exposure status.

An illustration from the above cohort study having Se = 0.9 and Sp = 0.9 is the following:

$$\text{True CIR} = \frac{150(.9) - 50}{150(.9) - 100}\,\frac{1000(.9) - 500}{1000(.9) - 500} = 2.43$$

The effect of equal misclassification of exposure status of diseased and nondiseased persons was to bias the RR estimate toward the null hypothesis of no effect, i.e., a bias from a true RR of 2.43 to an estimated RR of 2.0.

2. *Estimating the direction of bias caused by measurement error:*

a. *Nondifferential measurement errors:* If diseased and nondiseased persons are equally misclassified with respect to exposure status (Se is same for D and $\bar{D}$ and SP is same for D and $\bar{D}$), the estimate of effect will always be biased toward *the null hypothesis of no effect.*

Illustration for cumulative incidence ratios, nondifferential errors follows:

Measurement Error: Diseased		Nondiseased		Study Estimate of CIR	True CIR
Se	SP	Se	Sp		
0.9	0.9	0.9	0.9	2.0	2.43
0.7	0.7	0.7	0.7	1.68	2.43
0.7	0.9	0.7	0.9	1.62	2.43
0.9	0.7	0.9	0.7	1.77	2.43

Note: Lower sensitivity produces a larger bias than lower specificity of the same magnitude.

b. *Differential measurement errors:* If diseased and nondiseased persons are unequally misclassified with respect to exposure status (Se and/or Sp are not the same for D and $\bar{D}$), the measure of effect can be *biased toward or away from the null hypothesis.*

An illustration for differential Measurement Error:

Measurement Error: Diseased		Nondiseased		Study Estimate of CIR	True CIR
Se	Sp	Se	Sp		
0.9	0.9	0.9	0.7	1.38	2.43
0.7	0.9	0.9	0.7	0.82	2.43
0.9	0.7	0.9	0.9	2.56	2.43
0.9	0.7	0.7	0.9	3.51	2.43

Bias toward the null hypothesis occurs when:

(1) measurement errors are greater among nondiseased persons who are truly nonexposed but classified as exposed (row 1)

(2) measurement errors are greater among diseased persons who are truly exposed but classified as unexposed (row 2)

Bias is away from the null hypothesis when:

(1) measurement errors are greater among diseased persons who are truly nonexposed but classified as exposed (row 3)

(2) measurement errors are greater among nondiseased persons who are truly exposed but classified as nonexposed (row 4 which also includes low Sp for D group)

3. *Sensitivity and specificity applied to measurement of disease status:*

Observed cell frequencies in a cohort study

	D	$\overline{D}$	
E	a = 241	b = 2559	$N_1 = 2800$
$\overline{E}$	c = 158	d = 2042	$N_2 = 2200$
	$M_1 = 399$	$M_2 = 4601$	T = 5000

$$\text{Observed CIR} = \frac{241/2800}{158/2200} = 1.20$$

Assume nondifferential measurement errors in ascertainment of disease status:

$$Se = 0.866 \qquad Sp = 0.974$$

$$\text{True CIR} = \frac{[a - N_1(1 - Sp)]/N_1}{[c - N_2(1 - Sp)]/N_2}$$ For misclassification of disease status

$$\text{True CIR} = \frac{[241 - 2800(1 - .974)]/2800}{[158 - 2200(1 - .974)]/2200}$$

$$= 1.31$$

4. *To diminish information bias:*

a. Improve questionnaires and measuring instruments.

b. In data collection, pre-test questionnaires and train interviewers to be more objective and reproducible in their results.

c. In analysis, obtain information on sensitivity and specificity of measurements, so as to allow calculation of the direction of bias due to measurement error.

Bias Due to Confounding

Confounding is a distortion in the estimated measure of effect due to mixing of the study factor effect (exposure) with extraneous risk factor effects. Confounding variables are likely to occur in most observational studies, simply because most diseases are not only multifactorial in etiology, but their virulence or impact on a population can be considerably modified or even ablated by a variety of circumstances. For example, the infectivity of tubercle bacilli is altered by the racial composition, nutritional and socioeconomic status, and age of the host population, Similarly, asbestos appears to be a far more effective carcinogen for smokers than nonsmokers.

To be a confounder, a factor must possess the following characteristics:

1. The confounder must be an *independent* risk factor or effect modifier of the disease. The confounder must not be an intervening variable or link in a causal chain, as would be the case for smoking-induced metaplastic changes in bronchial lining cells, where smoking is the true independent risk factor, metaplastic change is the intervening variable, and lung cancer is the end point in the causal chain. Knowledge of the existence of risk factors or effect modifiers must come from the body of literature on the disease of interest.
2. The confounder must simultaneously be correlated with the distribution of the exposure factor.
3. The association (correlation) between confounder and exposure must be demonstrated in the study population. The confounding attribute of any risk factor is not an inherent association of risk factors in the population at large but is merely a relationship that happens to occur in the population selected for study. For example, there is no inherent association between being an asbestos worker and a cigarette smoker.

The confounding relationship can be schematically represented as follows:

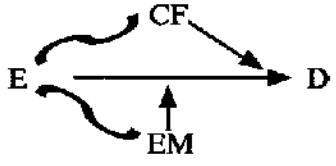
where the arrow indicates a causal relationship and the wavy line represents correlation but not

causality. The two essential features of a confounder are that it be an independent risk factor and that it be correlated in its distribution with exposure status.

Common examples of confounding factors that may be encountered in occupational health studies are:

1. Cigarette smoking as a potential confounder of the effect of occupational dust exposure or risk of chronic respiratory disease.
2. Alcohol habits as a potential confounder of the effect of exposure to an occupational liver carcinogen.
3. Dietary habits as a potential confounder of the effect of exposure to an assumed gastrointestinal carcinogen in the work or general environment.

Certain demographic characteristics of a population such as age, sex, and race are not biological "causes" of disease as such, but they alter or modify the apparent susceptibility of a population to disease. Many cancers and chronic degenerative diseases, such as emphysema and heart disease, are diseases of old age and are often more prevalent in males. Age and malehood modify the risk for these diseases in the sense that a population of older persons is at greater disease risk than one of younger persons. These *effect modifiers* (EM) can become confounding factors when their distribution is disproportionate between exposed and nonexposed groups. The complete schematic representation of confounding shows that confounding can result from the presence of an extraneous risk factor (CF) or an effect modifier (EM), either of which is differentially distributed between exposed and unexposed study groups.



In this diagram the partial arrow from EM to the E→D association is meant to signify that the relationship is different at one level of EM (young age) than at another (old age).

Example of confounding: a positive association between E and CF

Assume a case-control study of the association between surface sources of drinking water and colon cancer.

(1) Simple analysis

	D	$\overline{D}$	
E	170	80	250
$\overline{E}$	80	170	250
	250	250	500

$$OR = \frac{170(170)}{80(80)} = 4.52$$

(2) Stratified analysis by urban vs. rural residence

Urban

	D	$\overline{D}$	
E	150	30	180
$\overline{E}$	50	20	70
	200	50	250

$$OR = \frac{150(20)}{50(30)} = 2.0$$

Rural

	D	$\overline{D}$	
E	20	50	70
$\overline{E}$	30	150	180
	50	200	250

$$OR = \frac{20(150)}{50(30)} = 2.0$$

Note:

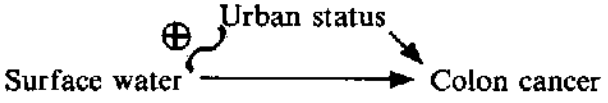
(a) In the rural stratum, 70/250 subjects are exposed to surface water. In the urban stratum, 180/250 are exposed to surface water. Thus, urban status is correlated with exposure to surface water.

(b) In the rural stratum, 50/250 subjects are

diseased. In the urban stratum, 200/250 subjects are diseased. The OR for disease, given urban vs. rural status is

$$\frac{200(200)}{50(50)} = 16$$

Thus urban status is a risk factor for disease. In this example, the measure of effect (the OR for surface water as a risk for colon cancer) in the simple analysis was confounded by urban status, which was both an independent risk factor for disease and was correlated with the distribution of the risk factor.



Example of confounding: a negative association between E and CF

Assume a cohort study of occupational dust exposure and chronic respiratory disease

	D	$\bar{D}$	
E	1,000	4,000	5,000
$\bar{E}$	1,000	4,000	5,000
			10,000

$$CIR = \frac{1{,}000/5{,}000}{1{,}000/5{,}000} = 1.0$$

(No apparent risk)

Stratify the population by smoking status

Nonsmokers

	D	$\bar{D}$	
E	350	3,650	4,000
$\bar{E}$	150	850	1,000

$$CIR = \frac{350/4{,}000}{150/1{,}000} = 0.58$$

Smokers

	D	$\bar{D}$	
E	650	350	1,000
$\bar{E}$	850	3,150	4,000

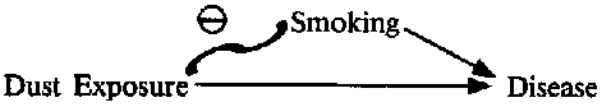
$$CIR = \frac{650/1{,}000}{850/4{,}000} = 3.06$$

Note that 4,000/5,000 nonsmokers are exposed to occupational dusts while 1,000/5,000 smokers are exposed.

Note also that smoking is a risk factor for disease (CIR = 3.0).

Thus, the true association of dust factor exposure with disease was confounded (in this case obliterated) by the negative correlation of smoking with dust exposure, when smoking itself was a risk factor for disease.

In this study, dust exposure had an effect on disease risk only in the presence of smoking. For smokers, dust exposure enhanced the risk of disease that was already increased by smoking alone.



The magnitude of the confounding effect can be simply quantified by the following equation (20):

$$RR_{CF} = \frac{RR_{Apparent}}{RR_{Standardized}}$$

where RR_{CF} is the relative risk (or other measures of effect) due to confounding, $RR_{Apparent}$ is the relative risk obtained when the confounding factor is not taken into account, and $RR_{Standardized}$ is the standardized relative risk measure obtained when the E→D relationship is adjusted for unequal distribution of the confounding factor between exposed and unexposed groups. $RR_{Standardized}$ can be obtained by computing SMR (or preferably an SRR) if age or some other single factor is responsible for con-

founding, or $RR_{Standardized}$ can be obtained by a stratified analysis that yields a summary estimate of overall effect adjusted for the distribution of several simultaneous confounding factors. Regression analysis or logistic risk functions can also be applied to an $RR_{Standardized}$.

Example

Assume a cohort study of 10,000 rubber workers followed for 10 years to evaluate the risk of benzidine exposure on bladder cancer incidence. Smoking is an independent risk factor for the disease and is correlated with the distribution of benzidine exposed workers.

Smokers

	D	Person-Yrs.
E	25	20,000
$\bar{E}$	15	30,000
	40	50,000

Nonsmokers

	D	Person-Yrs.
E	5	10,000
$\bar{E}$	10	40,000
	15	50,000

RR_A = Apparent RR due to benzidine

$$= \frac{\frac{25+5}{30{,}000}}{\frac{15+10}{70{,}000}} = 2.8$$

$RR_S = RR_{Standardized} = SMR$

$$= \frac{\text{Observed cases in exposed}}{\text{Expected cases in exposed}} =$$

$$\frac{25+5}{15/30{,}000(20{,}000)+10/40{,}000(10{,}000)} = 2.4$$

$$RR_{CF} = \frac{RR_A}{RR_S} = \frac{2.8}{2.4} = 1.17$$

We can conclude that the true RR due to benzidine exposure is 2.4 and that the higher apparent RR of 2.8 was due to the confounding effect of smoking which contributed 1.17 times the true RR to the apparent RR.

Methods for Controlling for Potential Confounders

Methods Used in the Selection of Subjects

1. Restricting of subjects to one category of the confounder or restricting eligibility into the study population for all subjects (e.g., only white males between the ages of 35 and 55 in 1960, or only nonsmokers).
2. Matching: restricting eligibility into the study population to subjects in the comparison groups (s) (e.g., pairing each case with one noncase of the same age, race, and sex). Matching along controls for confounding only in a cohort design. In a case-control design, matching must be coupled to a matched-pairs analysis to assure that confounding will be controlled.
   a. Individual matching—selecting one or more comparison subjects for each index subject so as to be similar with respect to one or more variables.
   b. Frequency matching—selecting a comparison group in such a way that it has the same distribution on one or more variables as does the index group.
3. Randomization—(in experiments) random allocation of "treatments" (i.e., the study factor) to the study population.
   a. "Simple" randomization—no consideration of other factors in the random allocation of treatments.
   b. "Restricted" randomization—consideration of other factors in the random allocation of treatments through blocking, grouping, and balancing.

Methods Used in the Analysis

1. *Stratification*—dividing the data into two or more extraneous variables, prior to further analysis (e.g., standardization). This is the main tool for ascertainment and control of confounding in epidemiologic analysis.
2. *Multivariate analysis*—using a statistical model to predict (or discriminate) the disease from two or more predictors, in-

cluding the study factor (e.g., multiple regression).

3. *Stochastic models*—fitting the data to a probabilistic model which assumes a particular configuration of factors, putatively involved in the etiology of a disease (e.g., Markov chain).

G. CRITERIA FOR INFERRING CAUSALITY

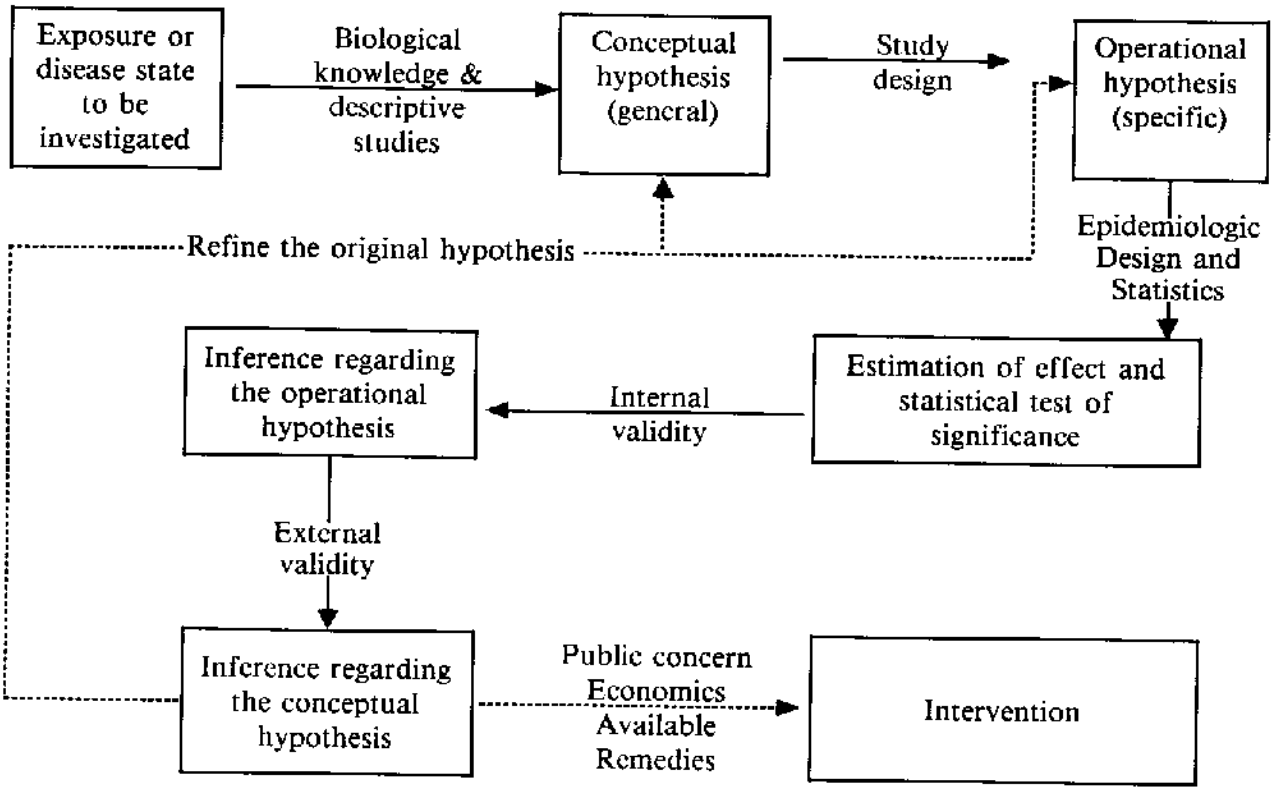
The process of inferring that an observed measure of effect (e.g., a relative risk of 2.5) implies causality entails answering three questions in sequence:

1. Is the effect (the relative risk of 2.5) a true effect in the sense that it is statistically significant, or is it merely a random observation, an extreme sample drawn from a population in which the true relative risk is 1.0? We answer the question by applying standard statistical methods with which we can measure the precision of our relative risk estimate.
2. Is the effect accounted for by something other than exposure, i.e., is the effect distorted by a systematic error, a bias due to selection, measurement errors, or confounding? To evaluate the possibility of bias, we must scrutinize the study design and the analysis and determine whether the investigators have avoided the various types of bias. We feel assured if the investigators use follow-up procedures for nonrespondents, measure sensitivity and specificity, and carefully examine the distribution of extraneous risk factors among exposed and unexposed groups. No study can be perfect in this regard, but we can attach a subjective weight to the evidence from each study as a function of the handling of potential biases.
3. Does the effect appear to be causal, i.e., is the exposure-disease association supported by evidence external to the study itself—by the total body of knowledge pertaining to the association between exposure and disease? Here we are referring to the external validity of the study, to the breadth of scientific generalization warranted by the addition of this study to the overall state of knowledge related to the study's conclusions. While formal statistical tests guide us in answering the first question, and epidemiological principles of design and analysis are helpful in answering the second, there is no organized methodology so far developed for approaching this question. We are forced to rely on educated judgments that are necessarily subjective, even though these judgments may be based upon commonly accepted rules of scientific inference. In epidemiology, scientific inferences concerning causality cannot yet be based on immutable laws, mathematical or statistical computations, or entirely objective and repeatable experiments. Considerable judgment, based upon the experience and wisdom of the judges, must be brought to bear in deciding whether a body of evidence warrants the conclusion that a true causal relationship exists.

The judgmental process follows the general scheme of reasoning illustrated in Table I-52 for epidemiological investigations. Epidemiologists usually begin with the need to evaluate some public health problem: a disease whose etiology is not fully explained by known risk factors or an exposure that may be hazardous to public health. Descriptive studies may be carried out, to provide clues regarding high risk groups, environments associated with excess disease, or temporal patterns of disease variation. More importantly, the epidemiologist must turn to other biological disciplines and to previous epidemiological investigations, to assess whether there is a biological basis for postulating an exposure-disease relationship. To proceed without this basis is to run the risk of generating spurious associations without causal implications. The conceptual hypothesis that evolves from this reasoning is a general statement concerning an exposure-disease association—e.g., beryllium exposure is a risk factor for lung cancer. The conceptual hypothesis is not tied to any source population. To evaluate the conceptual hypothesis, it is necessary to design a study that can test the conceptual hypothesis within the specific time, place, and person circumstances of a source population in which some or all of the population members are exposed to the study factor. At this point, the study hypothesis becomes operational, with specifications related to the size and composition of the study popula-

Table I-52

A PROCESS FOR DRAWING CAUSAL INFERENCE FROM EPIDEMIOLOGIC STUDIES



tion and the particular nature of its exposure, e.g., workers employed at a particular beryllium production process will show an excess relative risk of lung cancer when followed over the time period 1945-1975. It is now possible to choose an appropriate study design according to epidemiologic principles of good design; collect data; obtain measures of disease and exposure; estimate the effect; and apply tests to determine the statistical significance of the observed effect. The study now falls into the established framework of biostatistical analysis. Simultaneously, the study must be designed and analyzed to avoid the various forms of bias, and in interpreting their results, investigators need to consider whether other factors that could not be accounted for might have influenced the measure of effect. In most early studies, a careful scrutiny of results will reveal missing pieces of evidence, potential selection biases, inadequate measures of exposure, incomplete information on disease status, or inadequate data on other risk factors. From this evaluation, investigators are able to refine and often restrict their conceptual hypothesis or to reformulate an operational hypothesis that is now enriched with considerably more specificity. Progress in epidemiology, as in all of science, is made by finding the exceptions to the rule, discarding old and developing new hypotheses that better explain present and previous observations. A skeptical attitude toward his own results forces the investigator to rethink his conclusions, challenge his assumptions, and design fresh studies that may considerably strengthen the basic conceptual hypothesis.

The process of hypothesis testing, refinement of knowledge and retesting of hypothesis has no clear demarcation between evidence of firm association and of causation. By the nature of observational studies on human disease risks, we know that an association may be greatly

altered by circumstances of person, place, and time. The magnitude of disease risk in one plant may be entirely different in another, even though the same product is manufactured in both. We remain skeptical about the applicability of conclusions from one study until we see the results replicated by other investigators in other population groups. Even the first studies of cigarette smoking and lung cancer were greeted with healthy skepticism by well established scientists.

At some point, however, the state of knowledge is such that it is possible to review the range of studies and question whether the evidence is sufficient to infer causality. Such questions are frequently asked by federal agencies responsible for developing occupational and environmental health standards for public health protection.

In 1965, Austin Bradford Hill addressed the question of association or causation in a paper that has become a classic for its clarity and wide acceptance (9). Hill presents a series of criteria that can be considered in judging whether evidence for an association warrants a causal interpretation. These criteria, listed in Table I-53, are not, in the author's words, "indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do...is to help us to make up our minds on the fundamental question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more likely than cause and effect?"

All scientific evidence is incomplete, by the very nature of the hypothetical, deductive approach of the scientific method. A conclusion about causality may be upset or modified by advances in knowledge. However, it is unlikely a single study could contradict a body of evidence that meets the criteria of A. B. Hill. If a new hypothesis is advanced in competition with a well established conclusion, we should prefer the new hypothesis only if at least one of the following criteria is satisfied, as proposed by Buck in commenting on Karl Popper's philosophy of science (4):

1. The new hypothesis makes more precise predictions.
2. It explains more of the previous observations.
3. It explains the previous observations in more detail.
4. It has passed tests which the older hypothesis has failed.
5. It has suggested new tests or made new predictions not made by the older hypothesis.
6. It has unified or connected phenomena not previously considered to be related.

Table I-53

CRITERIA FOR INFERRING CAUSALITY IN EPIDEMIOLOGICAL STUDIES

1. Strength of the association (relative risk).
2. Consistency: replication of results by different investigators in different places, circumstances, and times.
3. Biological plausibility: depends on the current state of knowledge.
4. Biological coherence: agreement of results with findings of experimental research and clinical observations (coherence of evidence among experimental and observational disciplines.
5. Biological gradient: increase in disease with increase in intensity of exposure (dose-response curve).
6. Temporality: exposure precedes disease.
7. Specificity: the disease outcome is specific to, or characteristic of, exposure to a particular agent, e.g., pleural mesothelioma and asbestos (a weak criterion).
8. Effect of intervention: removal of putative cause results in significant reduction in disease incidence.
9. Analogy: drugs or chemicals that are structural analogues of a harmful agent may also induce similar harmful effects (a weak criterion).

Source: Hill, A.B.(9).

Hence, confidence in making a causal inference should not depend on the lack of any alternative explanation, but on the ability to consider many alternatives, all of which can be rejected.

REFERENCES

1. Archer, V.E., Wagoner, J.K., and Lundin, F.E.: Lung cancer among uranium miners in the United States. Health Phys *25:* 351-371, 1973.
2. Boice, J.D.: Follow-up methods to trace women treated for pulmonary tuberculosis, 1930-1954. Am J Epidemiol 107: 127-139, 1978.
3. Brinton, L. A., Blot, W.J., Stone, B.J., and Fraumeni, J.F., Jr.: A death certificate analysis of nasal cancer among furniture workers in North Carolina. Cancer Res 37:3473-3474, 1977.
4. Buck, C.: Popper's philosophy for epidemiologists. Int J Epidemio *4:*159-168, 1975.
5. Cole, P., Hoover, R., and Friedell, G. H.: Occupation and cancer of the lower urinary tract. Cancer *29:*1250-1260, 1972.
6. Copeland, K. T., Checkoway, H., Hollbrook, R. H., and McMichael, A. J.: Bias due to misclassification in the estimate of relative risk. Am J Epidemiol *105:*488-495, 1976.
7. Cornfield, J.: A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast, and cervix. J Natl Cancer Inst *11:*1269-1275, 1951.
8. .Enterline, P. E.: Mortality among asbestos products workers in the United States. Ann NY Acad Sci *132:*156-165, 1965.
9. Hill, A. B.: The environment and disease: association or causation? Proc Roy Soc Med *58:*295-300, 1965.
10. Horwitz, R. I. and Feinstein, A. R.: Alternative analytic methods for case-control studies of estrogens and endometrial cancer. New Engl J Med *299:*1089-1094, 1978.
11. Hutchinson, G. B. and Rothman, K. J.: Correcting a bias? (editorial). New Engl J Med *299:*1129-1130, 1978.
12. Lloyd, J. W. and Ciocco, A.: Long-term mortality study of steelworkers. I. Methodology. JOM *11:*299-310, 1969.
13. Lloyd, J. W.: Long-term mortality study of steelworkers. V. Respiratory cancer in coke plant workers. JOM *13:*53-68, 1971.
14. MacMahon, B. and Pugh, T. F.: Epidemiology. Principles and Methods. Boston, Little Brown, 241-282, 1970.
15. Mantel, N. and Haenszel, W.: Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst *22:*719-748, 1959.
16. McMichael, A. L., Spirtas, R., and Kupper, L. L.: An epidemiologic study of mortality within a cohort of rubber workers, 1967-72. JOM *16:*458-464, 1974.
17. McMichael, A. J.: Standardized mortality ratios and the "healthy worker effect": scratching beneath the surface. JOM *18:*165-168, 1976.
18. McMichael, A. J., Andjelkovich, D. A., and Tyroler, H. A.: Cancer mortality among rubber workers: an epidemiologic study. Ann NY Acad Sci *271:*125-137, 1976.
19. Miettinen, O. S.: Standardization of risk ratios. Am J Epidemiol *96:*383-388, 1972.
20 .Miettinen, O. S.: Components of the crude risk ratio. Am J Epidemiol *96:*168-172, 1972.
21. Morris, J. N.: Uses of Epidemiology. Third edition, Edinburgh, London, New York, Churchill Livingstone, 381 p, 1975.
22. Redmond, C. K. and Breslin, P. P.: Comparison of methods for assessing occupational hazards. JOM *17:*313-317, 1975.
23. Redmond, C. K., Strobino, B. R., and Cypress, R. H.: Cancer experience among coke by-product workers. Ann NY Acad Sci *271:*102-115, 1976.
24. Sackett, D. L.: Bias in analytic research. J Chronic Dis *32:*51-63, 1979.
25. Schlesselman, J. J.: Sample size requirements in cohort and case-control studies of disease. Am J Epidemiol *99:*381-384, 1974,
26. Shy, C. M., Kleinbaum, D. G., and Morgenstern, H.: The effect of misclassification of exposure status in epidemiological studies of air pollution health effects. Bull NY Acad Med *54:*1155-1165, 1978.
27. Walter, S. D.: Determination of significant relative risks and optimal sampling procedures in prospective and retrospective comparative studies of various sizes. Am J Epidemiol *105:*387-397, 1977.