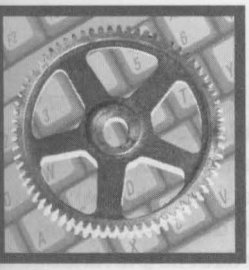


6 Epidemiology

Ellen A. Eisen and David H. Wegman

Epidemiologic studies further our understanding about environmental determinants of disease and serve as a basis for developing public health policy. All health professionals rely to some degree on epidemiologic literature. To remain current in their fields they need to be able to assess the quality of studies. This chapter is designed to assist health professionals in understanding how epidemiology is applied to occupational health and in critically interpreting the relevant epidemiologic literature.

The relation between any group of workers and their work environment is dynamic. New workers are hired and others leave the workforce. Exposures vary over time because of job transfers, changes in technology or production processes, and other factors. The workforce ages, and individual workers alter their personal habits, such as cigarette smoking. The epidemiologist uses analytic tools to examine this complex mix of variables in an attempt to understand the effects of workplace conditions on injury, disease, disability, and death.

Epidemiology complements clinical medicine in addressing occupational health problems. The clinical approach focuses on the individual and is concerned with diagnosis, treatment, and education of the worker regarding risk factors and preventive behavior. In contrast, the epidemiologic approach focuses on groups and is concerned with describing the distribution of injuries and disease in work groups, identifying population subgroups at high risk for a particular outcome, providing evidence for causal associations, estimating dose-response relations, and determining the effectiveness of preventive measures.

In epidemiology, the health outcome may be a discrete end point, such as diagnosis of a disease, or measurement of a biologic parameter, such as pulmonary function. The measure of exposure may be crude, such as membership in an occupational group, or more refined, such as the average daily time-weighted average exposure to a particular substance. Epidemiologists collaborate with toxicologists, ergonomists, environmental scientists, statisticians, and others to collect exposure data and to develop more precise methods for estimating exposures to chemical, physical, biomechanical, biologic, and psychosocial factors.

MEASURING EXPOSURE

Exposure is characterized by intensity or concentration, such as parts per million, and the duration over which it occurs. Cumulative or aggregate exposure is the product of intensity and duration and is an approximation of dose (to an organ or tissue). There are several degrees of refinement for approximating dose.

Potential Exposure

The most common available measure of exposure is simply a history of employment in

E. A. Eisen and D. H. Wegman: Department of Work Environment, University of Massachusetts Lowell, Lowell, Massachusetts 01854.

a specific industry or a specific job. Although this is a crude surrogate for exposure, if the relation between exposure and outcome is sufficiently strong a true association can be seen. For example, lung cancer was associated with asbestos in a study of shipyard workers, despite the fact that fewer than half of the shipyard workers had asbestos exposure (1). Nevertheless, the estimate of risk associated with exposure to a specific agent is greatly diluted by such a surrogate measure. A study of diesel exposure among railroad workers was largely negative, but only 7% of workers were found to have had exposure to diesel fumes.

Quantity of Exposure

Measures of exposure should include both intensity and duration. Because data on duration of employment may be more easily and accurately determined than intensity of exposure, duration is frequently used as the dose surrogate. It is often possible to document the number of years employed from payroll records or from union seniority records. Sometimes length of employment is unknown, but data such as pension plan eligibility may provide at least a dichotomous measure of duration (e.g., more than or less than 10 years of employment).

Exposure estimates are improved when industrial hygiene input is available, either as judgments of potential exposure or as measurements of actual exposure. Variation in exposure occurs as a result of changes in work assignments between days and within any given day, differences in work habits, seasonal changes in ventilation patterns, use of personal protective equipment, and other factors. Knowledge about these variations is used to adjust job-specific exposure estimates. Current exposure estimates alone can be used to study acute effects, but for the study of chronic effects such estimates must be integrated with job histories to develop a measure of cumulative exposure (2). A complete work history ideally includes documentation of time spent in specific jobs together with information on gaps in employment, such as prolonged sick leaves, periods of layoff, or military leaves.

Estimations of cumulative exposure ideally rely on compilation of current and historical industrial hygiene data and interviews of plant personnel about the history of changes in the production process and exposure controls. Estimates can be made of past exposures by reconstructing and testing old work environments. For example, in studies of pulmonary function in the Vermont granite industry, there was a need to account for old exposures, but no measurements were available. An old granite shed was reopened and operated without modern exhaust ventilation controls to arrive at appropriate estimates of the old exposures (3).

To compute cumulative exposure, estimated exposure levels are weighted by the number of years in successive jobs and summed over all jobs held by each worker. An implicit assumption in the computation of cumulative exposure is that 1 year of exposure to 20 fibers/cubic centimeters of asbestos is equivalent to 10 years of exposure to 2 fibers/cc. Furthermore, exposure that occurred years ago is assumed to be biologically equivalent to the exposure last year. More complex weighting schemes are possible but should be based on specific biologic hypotheses about the relative importance of different exposure patterns. For example, exposures in the distant past can be weighted more heavily than those in the recent past for diseases such as silicosis, in which irreversible changes are believed to accumulate gradually over years.

Biologic Monitoring

Evaluation of workers for toxic agents (or their metabolites) in blood, urine, or exhaled air sometimes permits improved estimation of real dose. One advantage of a biologic index is that it accounts for exposures from multiple routes of absorption, including inhalation, skin absorption, and ingestion. For example, urinary hippuric acid levels can be used to estimate total recent dose of toluene

to an exposed worker via both inhalation and skin absorption. Another advantage of biologic markers is that they may reflect exposure over specific time intervals. For example, although blood lead levels indicate recent lead exposure, x-ray fluorescence of bone provides an estimate of total body burden of lead, reflecting long-term exposure (4). Although no biologic monitoring tests currently exist for a substantial number of hazardous workplace substances, biologic monitoring is receiving more attention today and new measures of the body burden of toxic agents can be expected.

In summary, there are a variety of ways to estimate both current and past exposures. An accurate measurement of exposure is equally as important as an accurate measurement of health outcome in arriving at an unbiased and precise estimate of the exposure-outcome relationship.

COMMON MEASURES OF DISEASE FREQUENCY

If a disease is extremely rare, the occurrence of even a few cases can prompt further investigation of a possible workplace hazard. For example, three cases of angiosarcoma diagnosed during a 3-year period among a group of workers exposed to vinyl chloride was sufficient to make a plant physician suspect that the chemical was a carcinogen (5). In most instances, however, a count of cases cannot be interpreted without knowing the size of the population from which the affected workers came. The problem is illustrated by the example of a study of workers in a coated fabrics plant, 68 of whom were found to have a peripheral neuropathy (6). Even though this end point is uncommon in the general population, it is not sufficiently rare that the expected number of cases can be treated as zero. A case count by itself has little or no meaning without a standard of reference; that is, disease frequency can be interpreted only in relation to the size of the population at risk. The measures of disease frequency that are most commonly used are prevalence and incidence.

Prevalence

The simplest quantity, known as *point prevalence,* is the ratio between the number of cases present and the size of the population at risk *at a single point in time.*

Point prevalence
= No. cases ÷ Total population

To interpret the public health significance of the 68 cases of peripheral neuropathy in the coated fabrics plant, we first need a denominator. The total plant population was 1,157, so the point prevalence was 68 ÷ 1157 = 5.9%. To determine whether this is excessive, the prevalence in the plant must be compared with the prevalence in the general population or in another, more appropriate comparison group. A limitation of point prevalence is that it does not distinguish between old and new cases.

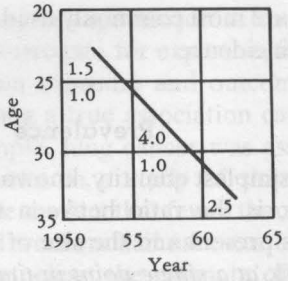
Incidence Rate

By contrast, *incidence* measures the occurrence of new cases. The incidence rate is based on the number of new cases occurring *during a specified period of time.*

Incidence rate
= No. new cases ÷ Total population at risk

In the coated fabrics plant, only 50 affected workers had onset of the disease within the past year; 18 of the 68 prevalent cases occurred more than 1 year ago. Therefore, the population at risk for development of a new case within the past year was 1,157 − 18 = 1,139. Because the number of new cases in that period was 50, the plant-wide annual incidence rate was 50 ÷ 1,139 = 4.4% for the preceding year.

The incidence rate can also be refined to reflect monitoring of individual workers for varying lengths of time. The appropriate denominator incorporates the concept of *person-time*, usually expressed in units of per-



Time Period	Age Group	Person-years
1950 – 1954	20 – 24	1.5
	25 – 29	1.0
1955 – 1959	25 – 29	4.0
	30 – 34	1.0
1960 – 1964	30 – 34	2.5
		10.0

FIG. 6-1. Person-years experienced by a worker entering a follow-up program at age 23 years 6 months in mid-1952 and leaving in mid-1962. (Adapted from Monson RR. Occupational epidemiology, 2nd ed. Boca Raton, FL: CRC Press, 1989.)

son-years. This denominator takes into account not only the number of at-risk persons but also the period during which they were at risk for development of the disease. An example of how to calculate the contribution of a single worker to a person-years denominator is illustrated in Fig. 6-1.

COMPARISONS OF RATES

To understand whether an incidence rate in an exposed population is excessive, it is necessary to compare it with the rate in an unexposed population. The two most common comparisons, or estimates of risk, are *relative risk* (the ratio of rates) and *attributable risk* (the difference between rates).

Relative Risk

The relative risk, or rate ratio, is designed to communicate the relative importance of an exposure by comparing rates from an exposed population with those from a nonexposed or normal population. In its simplest form, it is the ratio of two rates (Table 6-1). In the case of the fabrics plant, the suspect neurotoxin was in the print department, so it was possible to create a within-plant comparison. Of the 1,139 disease-free workers in the plant, 169 worked in the print department and 34 of those had onset of peripheral neuropathy in the past year, resulting in an annual incidence rate of 34 ÷ 169 = 20.1%. Among the remaining 970 workers, there were 16 new cases, resulting in an annual incidence rate of 16 ÷ 970 = 1.6%. The relative risk (or *incidence rate ratio*) therefore was .201 ÷ .016 = 12.6.

When examining different diseases or the effects of different hazards, relative risks can be compared directly. For example, the relative risk of lung cancer in heavy smokers compared with nonsmokers is very large (32.4), whereas that for cardiovascular disease is small (1.4). This suggests that smoking is more potent as a lung carcinogen than as a cardiotoxic agent.

Attributable Risk

Whereas the relative risk is a measure of the potency of the hazard, attributable risk measures the magnitude of the disease bur-

TABLE 6-1 *Derivation of relative and attributable risk**

	Exposure		
Disease	Present	Absent	Total
Present	a	c	a + c
Absent	b	d	b + d
Total	a + b	c + d	a + b + c + d

* Calculations:
Exposed disease rate = a/(a + b)
Nonexposed disease rate = c/(c + d)
Relative risk = a/(a + b) ÷ c/(c + d)
Attributable risk = a/(a + b) − c/(c + d)

den in the population ascribed to the exposure under study. This concept is particularly useful in occupational disease studies because occupational exposure is generally only one of several possible causes of any specific disease. The attributable risk is calculated by subtracting the rate of the particular disease in the nonexposed population from that in the exposed population (see Table 6-1). This risk difference is attributed to the exposure. In the coated fabrics plant, the incidence rate in the unexposed population (.016) is subtracted from the rate in the exposed population (.20), yielding an attributable risk of .184.

In the example of the impact of cigarette smoking on health, Table 6-2 shows that the smoking-attributable risk for lung cancer (2.20/1,000) is smaller than the smoking-attributable risk for cardiovascular disease (2.61/1,000). The attributable risk takes account of both the potency and the magnitude of the disease in the population. Despite the lower relative risk of cardiovascular disease due to smoking, the larger attributable risk indicates that reduction of smoking has a greater impact on cardiovascular disease than on lung cancer in a population.

Relative risks are commonly presented in epidemiologic studies as a measure of association between an exposure and a disease outcome. In contrast, attributable risks are useful in setting priorities for public health interventions or control.

INTERPRETING RATES

Crude Rates. When rates are calculated without consideration of factors such as age or calendar year, they are referred to as *crude* rates. Crude rates can be misleading. For example, if the exposed group includes a high proportion of elderly persons and disease incidence increases with age, then observed differences in crude rates may only reflect differences in age.

Specific Rates. These are rates estimated for homogeneous subgroups of a population defined by specific levels of a factor, such as age-specific rates. Sometimes an elevated disease risk exists only in one subgroup.

Adjusted Rates. Although specific rates can sometimes provide valuable information, it is cumbersome to compare many specific rates. Methods have been developed for estimating a single summary rate that takes account of differences in the distribution of population characteristics such as age. Such rates are known as *adjusted* or *standardized* rates. Two types of adjustment are commonly used: *direct* adjustment (rates in the study population are weighted by person-time in a reference population) and *indirect* adjustment (rates in a reference population are weighted by person-time in the study population). These methods can be illustrated with examples of adjustment for age (Table 6-3). For a description of these types of adjustment, see the appendix at the end of the chapter.

TABLE 6-2 *Relative and attributable risk of death among British male physicians from selected causes associated with heavy cigarette smoking*

Cause of death	Death rate*		Relative risk	Attributable risk
	Nonsmokers	Heavy smokers†		
Lung cancer	0.07	2.27	32.4	2.20
Other cancers	1.91	2.59	1.4	0.68
Chronic bronchitis	0.05	1.06	21.2	1.01
Cardiovascular disease	7.32	9.93	1.4	2.61
All causes	12.06	19.67	1.6	7.61

*Number of deaths per 1,000 per year.
†Smokers of ≥25 cigarettes per day
From Doll R, Hill AB. Mortality in relation to smoking: ten years' observations of British doctors. Br Med J 1964;1:1399.

TABLE 6-3 *Age effect on incidence of myocardial infarction**

Location	Workers <45 yr			Workers ≥45 yr			All workers			
	Cases	Population at risk	Age-specific incidence rate	Cases	Population at risk	Age-specific incidence rate	Cases	Population at risk	Crude incidence rate	Age-adjusted incidence rate†
Factory 1	4	400	10.0	18	600	30.0	22	1,000	22.0	18.0
Factory 2	10	800	12.5	10	200	50.0	20	1,000	20.0	27.5

* The incidence rate is expressed as new myocardial infarctions occurring in a 10-year period of observation per 1,000 population.
† Based on age distribution summed for factory 1 and factory 2.

Drawing by Nick Thorkelson.

TYPES OF EPIDEMIOLOGIC STUDY DESIGNS

Epidemiologic studies can be categorized into three general types: cohort, case-control, and a hybrid called cross-sectional. The population in a cohort study is defined on the basis of exposure status and often represents a complete enumeration of both current employees and past workers. The cohort is monitored over time, and the incidence of symptoms, functional abnormalities, disease, and death are observed. By contrast, the study group in a case-control study is defined on the basis of health status (Fig. 6-2), and exposures are compared between subjects with and without disease. The cross-sectional design typically focuses on active employees at a single point in time, collecting both exposure and health information simultaneously.

Cross-Sectional Studies

The cross-sectional approach is commonly used in field investigations because it is the simplest study design to execute. Either the prevalence of disease is compared between groups defined by exposure status, or the prevalence of exposure is compared between groups defined by disease status. Exposure can be classified dichotomously (e.g., exposed versus nonexposed) or along a gradient (e.g., high, medium, and low). Exposure classification can be based on either current or lifetime exposure.

Example: Cross-sectional Study, Exposure-Based

A pathology resident died of an acute heart attack at the age of 28 years. In discussing this incident, a number of the other pathology residents noted that they had been experiencing abnormal heart rhythms (palpitations). Those with palpitations had all worked with fluorocarbon propellants, which were used to prepare frozen sections of pathology tissue and to clean instruments or specimen slides. This discovery led to a study of all pathology department employees (the exposed group). Employees of a radiology department of similar size and distribution of physicians and nonphysician staff members were selected as a nonexposed comparison group. Each person was asked about occur-

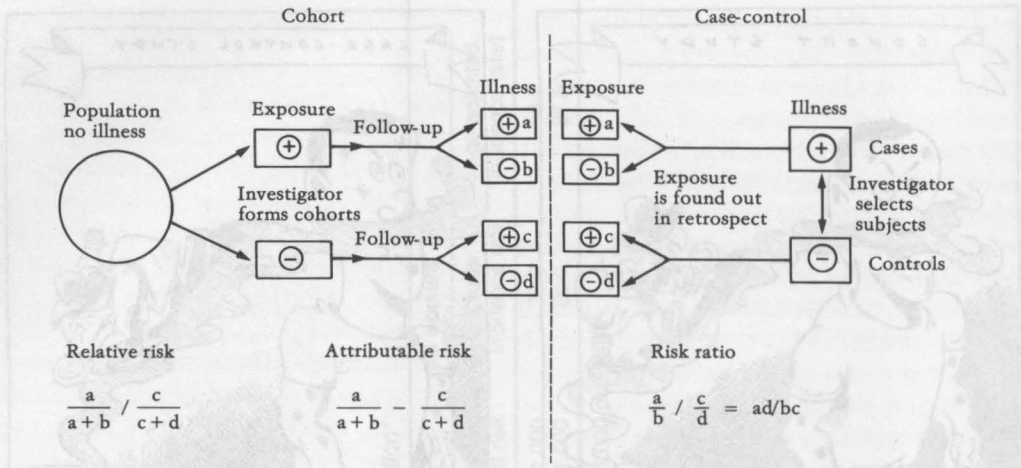


FIG. 6-2. General outline of cohort and case-control studies.

rence of palpitations and current use of fluorocarbon propellants. Those exposed to the propellants had twice the prevalence of palpitations as those not exposed (7).

Example: Cross-sectional Study, Disease-Based

In the study just described, the investigators wanted to further explore their hypothesis that exposure to fluorocarbon propellants accounted for the elevated risk of palpitations among pathology department employees. A follow-up analysis was designed in which groups were defined on the basis of disease prevalence (7). The pathology department staff was divided into those with palpitations (cases) and those without palpitations (controls). Forty percent of the cases but only 20% of the controls had exposure to fluorocarbon propellants, for an odds ratio (OR) of 2.7. The OR is an approximation of the relative risk (see later discussion).

Cohort Studies

In a cohort study design, an exposed group is identified and monitored forward in time to measure the occurrence of adverse health outcomes. The incidence is observed in the study group and compared with that in a nonexposed, reference group. Cohort studies are described as *retrospective* (the cohort is defined at some point in the past and monitored to the present) or *prospective* (the cohort is defined at the present and monitored into the future).

Cohort Mortality Studies. Although the cohort design can be used to examine nonfatal outcomes, most occupational cohort studies examine mortality from specific causes. The most common of type of cohort study is the standardized mortality study, in which the cause-specific mortality rate of the exposed cohort is compared with that of the general population (assumed to be nonexposed). This comparison results in an approximation of relative risk, known as the *standardized mortality ratio* (SMR). If the number of deaths observed in the exposed cohort is equal to the number expected based on death rates in the standard population, the SMR equals 1.0, which indicates neither an excess nor a deficit of risk. If the SMR is greater than 1.0, the data suggest an increased risk in the exposed population.

To conduct an SMR study, the following information must be obtained for each member of the cohort: date of birth, date of entry into cohort, date of leaving cohort, vital status (alive or dead), and cause of death for those who died. With these data, person-

years at risk can be determined that take into consideration times when workers entered or left during the study period. This permits a calculation of years at risk, adjusting for length of time since entry into the study and for age. This type of study requires personnel records with accurate employment data; if such data on the total population at risk are lacking, the mortality experience can be evaluated by proportional mortality analysis.

Cause-specific *proportional mortality ratios* (PMRs) are calculated as the proportion of all deaths attributed to each specific cause of death. These ratios in the study population are compared with those from the general population and are adjusted for age, sex, race, and year of death (again, indirect standardization). In contrast to an SMR study, an excess or deficit of deaths from causes other than the one under scrutiny in a PMR study can affect the proportional distributions. Therefore, PMR study results are less reliable than SMR study results.

Example: Retrospective Cohort Mortality Study, Standardized Mortality Ratio

A study of mortality in steelworkers was planned in 1962 (8). The workers were selected for study if they were employed in 1953; they were monitored until the end of 1962 for vital status. More than 59,000 workers participated, and 4,716 deaths recorded. When the numbers of deaths from specific causes were compared with the numbers expected, based on deaths in the study county, there appeared to be no excesses. (SMR for all cancers combined was only .92.) The study population was large enough, however, to examine the SMRs for particular subgroups and compare them in each case with the SMR for the rest of the workers (documented as nonexposed). This in-depth evaluation led to the discovery that lung cancer risk appeared to be higher among the coke plant workers and that the excess was much greater for the nonwhite employees. Further analysis showed that those who worked on top of the coke ovens (the most heavily exposed job assignment) had the highest risk for lung cancer (Table 6-4). The large size of the study population permitted detailed examination of a number of subgroups. As a result, the very high risk of lung cancer in coke oven workers was extracted from the overall unremarkable results.

TABLE 6-4 *Standardized mortality ratios for lung cancer among nonwhite male coke oven workers for ≥5 yr*

Work area	Observed	Expected	SMR
Side oven	19	9.5	2.00
Part-time topside	9	2.3	3.91
Full-time topside	23	1.9	12.10
Total	51	19.0	2.68

Adapted from Redmond CK, Wieand HS, Rockette HE, et al. Long-term mortality study of steelworkers. Department of Health and Human Services (NIOSH) Publication No. 81-120, U.S. Government Printing Office, Washington, DC, 1981.

Cohort Morbidity Studies. Increasingly, the cohort design is being used to study occupational risks associated with a variety of nonfatal health outcomes. Retrospective studies can be conducted if information on past health status is available (e.g., in medical records or collected in health surveys). More often, morbidity studies require prospective study designs so that the health information can be collected directly by administering medical examinations, physiologic tests, or surveys of current health status. Studies that look at episodic health events, such as recurrent symptoms or changes in pulmonary function, are referred to as *longitudinal* studies, and the change in health status over time becomes the outcome.

Example: Retrospective Cohort Morbidity Study

A cohort of approximately 1,000 hospital nurses was studied to examine possible reproductive effects associated with use of sterilizing agents (9). Questionnaires and medical records were used to collect information retrospectively about both exposure and pregnancy history as far back as 30 years.

The frequency of spontaneous abortion among nurses currently using the sterilizing agents was only slightly higher than that for currently nonexposed nurses. A more striking difference was observed when results were stratified according to whether exposure to sterilizing agents had occurred during a past pregnancy. Among those exposed, the rate of spontaneous abortion was 16%, compared with 6% among the nonexposed. Of the three specific sterilizing agents considered, ethylene oxide showed the strongest association with spontaneous abortion.

Example: Prospective Cohort Morbidity Study

A study was designed to characterize adverse respiratory effects associated with exposure to toluene diisocyanate (TDI). Because TDI was already known to be a cause of asthma, the study was designed to measure other types of acute and chronic respiratory effects. Pulmonary function tests (including measurement of the forced expiratory volume in 1 second, FEV_1) were administered to all workers at a polyurethane-foam manufacturing firm on the first shift on a Monday morning. The workers were then divided into three exposure groups and retested at the end of the workday—a 1-day, prospective study design.

Generally, FEV_1 changes only slightly over the course of a workday. However, on average these workers were losing lung function over the duration of the shift, and the amount of loss increased with exposure (Table 6-5). To examine whether this acute response (presumably caused by bronchospasm) reflected a more persistent chronic effect, those who were still employed were retested 4 years later. The annual decline in FEV_1 was estimated for the same three exposure groups (10). Again, an exposure-related response was observed (see Table 6-5), with the high-exposure group losing lung function at the greatest annual rate. Cigarette smoking habits did not explain the effects noted in either the 1-day or the 4-year prospective study.

TABLE 6-5 *Acute and chronic change in FEV_1, by exposure group, in polyurethane foam manufacturing workers**

Exposure group	FEV_1 differences from beginning to end of study period	
	Acute (1-day) change (ml)	Chronic change over 4 yr (ml/yr)
Low	−78	0.5
Medium	−108	−33.3
High	−180	−60.5

* Negative change means loss over time.

Adapted from Wegman DH, Musk AW, Main DM, et al. Accelerated loss of FEV-1 in polyurethane production workers: a four year prospective study. Am J Indus Med 1982;3:209–215.

Case-Control Studies

In the case-control, or case-referent, study design, the investigator compares the frequency of exposure between groups with and without the disease of interest (see Fig. 6-2). The case-control design is particularly well suited to study diseases that occur with low incidence; a cohort study would have to be prohibitively large to generate enough cases to study.

There are three types of case-control studies: (a) studies nested within occupational cohort studies, (b) population-based case-control studies, and (c) registry-based case-control studies. In nested case-control studies, all cases of the selected disease are identified from the cohort, and controls are sampled from among those without the disease. In a mortality study, disease status may be determined at death from a particular disease; in a morbidity study, disease status may be determined by disease incidence based on diagnosis. In a population-based case-control study, all cases occurring in residents of a defined geographic area are included, and controls are selected from the same defined population. In a registry-based case-control study, cases of disease that are reported to

the registry with onset during a defined time period are identified, and controls are selected from the same registry base. Because registries often are not population-based (e.g., a hospital cancer registry), the selection of controls requires identification of patients with other diseases from the same source as the cases (e.g., from the same hospital).

The measure of risk typically calculated in a case-control study is the OR. The OR is a ratio of the odds of exposure (exposed to non-exposed) among the cases compared with the odds of exposure among the controls. From Table 6-1, it can be seen that a/b is the odds of exposure among the cases, and c/d is the odds of exposure among the controls. ORs approximate the incidence rate ratios that are obtained in cohort studies. Their interpretations are similar: OR = 1 means that there is no excess or deficit of risk.

A case-control study need not include all the cases within a defined population. Valid results may still be obtained when the case group includes only a sample of all cases. The major requirement for a valid case-control study is that the controls selected be comparable to the population from which cases were identified and that both cases and controls be selected without prior knowledge of past exposure history.

Example: Population-Based Case-Control Study

Non-Hodgkin's lymphoma has been associated with agricultural pesticide use in men, but little is known about risks in women. To address this lack of knowledge, National Cancer Institute investigators conducted a population-based case-control study (11) in which cases were defined as incident cases of non-Hodgkin's lymphoma among women residing in 66 counties in eastern Nebraska, diagnosed between 1983 and 1986 in all area hospitals. Controls were selected from female residents in the same counties using random digit dialing. No risk was found to be related to living or working on a farm. Small risks were observed for women who personally handled insecticides (OR = 1.3) or herbicides (OR = 1.2), and women who personally handled organophosphate insecticides had a 4.5-fold increased risk (Table 6-6). Because non-Hodgkin's lymphoma is a rare disease with a long latency, the case-control design was more feasible than a cohort study. Because exposures occur on farms, each of which employs a small number of workers, a community-based study was more practical than a workplace-based study.

Example: Nested Case-Control Study

The carcinogenic risk of pulsed electromagnetic fields was studied in a series of case-control studies nested in a cohort of electric utility workers (12). Case groups were defined as all diagnosed cases of selected cancers that occurred at any time after entry into the cohort until the end of follow-up in 1988. Controls were chosen at random from sets of cohort members matched to each case who had survived to the date of diagnosis of the case. Cumulative exposures were estimated up to the date of diagnosis of the case. Smoking information was obtained from company medical records. No associations were found between exposure to pulsed electromagnetic fields and cancers previously suspected of being associated with magnetic fields. However, the investigators reported a clear association between cumulative exposure to pulsed electromagnetic fields and lung cancer (after adjusting for cigarette smoking history), with an OR of 3.1 in the highest exposure category.

SELECTION OF TYPE OF STUDY

The study designs described have relative strengths and weaknesses. The choice of design is based on a variety of factors.

Cross-Sectional Studies

Cross-sectional studies have several advantages over cohort studies. First, cross-

TABLE 6-6 *Non-Hodgkin's lymphoma according to insecticide use among women in eastern Nebraska*

Insecticide class	Used on farms			Personally handled		
	Cases	OR	95% CI	Cases	OR	95% CI
Any insecticide	56	0.8	0.5–1.3	22	1.3	0.7–2.3
Chlorinated hydrocarbons	20	1.6	0.8–3.1	5	1.7	0.5–5.8
Organophosphates	14	1.2	0.6–2.5	6	4.5	1.1–17.9
Metals	3	1.6	0.3–7.5	0	—	—

OR, odds ratio; 95% CI, 95% confidence interval.

Adapted from Zahm SH, Weisenburger DD, Saal RC, Vaught JB, Babbitt PA, Blair A. The role of agricultural pesticide use in the development of non-Hodgkin's lymphoma in women. Arch Environ Health 1993;48:353–358.

sectional studies permit the examination of disease morbidity or measures of physiologic function. Second, because the subjects are alive at the time of the study, it is often possible to collect information directly on nonoccupational risk factors such as cigarette smoking or diet (potential confounders). Finally, because both disease prevalence and exposure data are collected at one point in time, cross-sectional studies usually require less time to complete than cohort or case-control studies.

These studies also have important limitations. They are regarded as less appropriate for investigating causal relations because they are based on prevalent, rather than incident, cases of disease. A second limitation is that they are based on actively employed workers and do not include employees who retired or terminated their employment before the beginning of the study. In the presence of an occupational hazard, workers whose health has been affected are more likely to leave the workforce; therefore, the absence of such workers may result in an underestimate of the association of interest.

Cohort Studies

These studies focus on exposure and look ahead to outcome or disease incidence. Several outcomes can be studied in the same population. Data collected on exposure retrospectively depends on the quality of past records, in contrast to data collected prospectively according to a specific study plan. If questionnaires or interviews about past exposures are used, selective recall can be a source of bias. Furthermore, because retrospective studies typically rely on outcomes recorded for other purposes, the end point is more likely to be cause of death than an earlier marker of disease. In comparison to cross-sectional studies, of cohort studies have the advantage of including the entire population of interest. However, the difficulty of long-term follow-up means that some subjects are inevitably "lost."

Case-Control Studies

The principal advantage of the case-control study is its relative simplicity and relatively low cost. Case-control studies are valuable when multiple exposures are being explored in the etiology of a disease. If the investigator wishes to examine a spectrum of diseases associated with an exposure, such as lead, a cohort study is desirable; but if the interest is in the causes of a specific disease, such as bladder cancer, then the case-control study is more suitable.

Case-control studies are regarded as slightly more susceptible to biases than cohort studies (see below). For example, exposure information may be recalled differently by subjects with and without disease. Moreover, there is the need to identify a control group with the same general exposure history of the population that generated the cases.

PROBLEMS RELATED TO VALIDITY

Because epidemiologic studies are observational studies rather than randomized experi-

ments, they are prone to biases, some of which are unavoidable. Careful consideration needs to be given to a study's validity (lack of bias). *Bias* is defined as a distortion of the measure of association between exposure and health outcome, such as an SMR or an OR. The degree to which the inferences drawn from a study are warranted is determined largely by the absence of bias. Reports of epidemiologic studies should provide sufficient information for the reader to understand what potential sources of bias were present and how these biases were addressed. There are three sources of bias: selection, misclassification, and confounding.

Selection Bias

Selection bias results from the inappropriate inclusion or exclusion of subjects in the study population. For example, in the past it was customary in studies of pulmonary function to exclude subjects who did not perform reproducible pulmonary function tests. It was subsequently discovered that subjects who had difficulty performing a reproducible forced expiratory maneuver had compromised respiratory health (13). The exclusion of such subjects could result in an overestimation of the respiratory health of a working population and possibly the underestimation of a dose-response association, if one exists.

Most types of selection bias, such as exclusion of short-term workers from the study population, cannot easily be corrected or controlled for in the analysis; they can only be prevented. To prevent selection bias in a cohort study, investigators should be kept unaware ("blinded") of cohort members' outcome status. Similarly, in case-control studies, investigators should be blinded as to the exposures status of cases and controls. Furthermore, selection of subjects should not be influenced by prior knowledge or suspicion of health outcome in a cohort study or of exposure status in a case-control study.

The most common type of selection bias in occupational epidemiologic studies is the "healthy worker effect" (HWE). This bias results from workers' selecting themselves out of the study groups rather than investigator error or oversight. For example, as described earlier, cross-sectional studies may result in an underestimate of the dose-response association if the occupational exposure causes disease which, in turn, causes workers to leave the workforce. Another example of HWE, common to cohort mortality studies, occurs because employed people are healthier than the general population, which includes the aged, the chronically ill, and those who are otherwise unfit to obtain and maintain employment.

As a result of the HWE, studies of illness or death among working populations often show lower rates of chronic diseases (e.g., cardiovascular diseases) than in the general population. In the mortality study of steelworkers described previously, the overall SMR, expected to be 1.00, was only .82 when the mortality rates of the surrounding county were used for comparison.

It is rare that an appropriate alternative comparison group of sufficient size is available. When possible, HWE bias is minimized by using a nonexposed comparison group drawn from within the study population. The HWE is reduced, although not necessarily eliminated, when analyses are based on this sort of "internal" comparison between exposed and nonexposed workers.

Misclassification

Misclassification (information bias) refers to an investigator's inadvertent placement of a worker into an incorrect category or group. Either disease or exposure can be misclassified. However, for purposes of illustration, the focus here is on exposure misclassification, because it is more relevant for occupational epidemiology. There are two types of misclassification: nondifferential and differential.

Exposure misclassification that is *nondifferential* is random misassignment of exposure that occurs regardless of disease status. Nondifferential misclassification is common

in occupational studies, in which there is often little information on subjects' exposures and subjects cannot be well classified into exposure categories. The problem is generally worse in retrospective studies, because adequate documentation of historical exposures is more difficult. The net effect is to reduce a study's ability to detect exposure-disease associations when such associations truly exist. Thus, nondifferential misclassification may result in an existing occupational hazard going unrecognized.

Bias of a different sort is presented by *differential* misclassification, in which the likelihood of misassignment of exposure is related to disease status. This type of bias can result in either a stronger or a weaker association than truly exists. In cohort studies, differential misclassification is commonly prevented by keeping the investigators blinded to exposure status during collection of outcome information. In this manner, any errors in collection should be randomly distributed among both exposed and nonexposed groups. In case-control studies, control of differential bias is much more complicated; it is difficult for the investigator, and usually impossible for the subject, to be unaware of the disease status when exposure information is being obtained. Prevention of differential misclassification depends on collecting data as objectively as possible.

Confounding

Confounding is present when two study groups (e.g., exposed and nonexposed) are not comparable with respect to a characteristic that is *also* a risk factor for the disease. For example, in a study comparing stomach cancer in coal miners and iron miners, chewing tobacco was considered to be a potential confounder because (a) it is used more commonly by coal miners who are prohibited from smoking in coal mines and, (b) it may be an independent risk factor for stomach cancer.

Confounding can be controlled either in the design of the study or in the analysis of the data. In case-control studies, matching of study subjects on potential confounders in the design phase can facilitate control of confounding in the analysis. To control confounding in the example of the stomach cancer study, subjects could be matched on tobacco-chewing habits so that the proportion of tobacco chewers is the same among cases and controls.

Stratification is the major approach to control of confounding in the analysis phase of cross-sectional, cohort, and case-control studies. A confounder such as age is used to define strata (e.g., 10-year age groups). The exposure-response association is then estimated in each stratum. Stratification, however, becomes problematic as the number of confounders increases, because the strata become too small to allow stable measures of risk. For example, if age, smoking, race, and gender must be controlled for simultaneously, there may be no nonsmoking 40-to-45-year-old white females in the study population. In this case, stratification becomes an inadequate method of controlling confounding, and mathematical modeling must be used to control confounding statistically.

Multivariate models impose particular mathematical forms on the dose-response relations, such as a linear or exponential form. By restricting the data to a specific structure, one can interpolate between sparse strata. Mathematical modeling generally involves "smoothing" of the data the distributions of confounders and exposure categories.

INTERPRETATION OF EPIDEMIOLOGIC STUDIES

The interpretation of epidemiologic studies depends on the strength of the association, the validity of the observed association, and supporting evidence for causality (Box 6-1). The strength of an association usually is measured by the size of the relative risk in studies of discrete health outcomes such as cancer, or by the magnitude of the difference between groups in studies of physiologic pa-

Box 6-1. Guide for Evaluating Epidemiologic Studies

To assist health professionals in reading, understanding, and critically evaluating epidemiologic studies, the following questions, adapted from Monson's *Occupational Epidemiology* (15), should serve as a useful guide.

Collection of Data

1. What were the *objectives* of the study? What was the association of interest?
2. What was the primary *outcome* of interest? Was it accurately measured?
3. What was the primary *exposure* of interest? Was it accurately measured?
4. What *type of study* was conducted?
5. What was the *study base*? Consider the process of subject selection and sample size.
6. *Selection bias:* Was subject selection based on the outcome or the exposure of interest? Could the selection have differed with respect to other factors of interest? Were these likely to have introduced a substantial bias?
7. *Misclassification:* Was subject assignment to exposure or disease categories accurate? Were possible misassignments equally likely for all groups? Were these likely to have introduced a substantial bias?
8. *Confounding:* What provisions, such as study design and subject restrictions, were made to minimize the influence of external factors before analysis of the data?

Analysis of the Data

9. What *methods* were used to control for confounding bias?
10. What *measure of association* was reported in the study? Was this appropriate?
11. How was the *stability* of the measure of association reported in the study?

Interpretation of Data

12. What was the *major result* of the study?
13. How was the interpretation of this result affected by the previously noted *biases*?
14. How was the interpretation affected by any nondifferential *misclassification*?
15. To what larger population may the results of this study be *generalized*?
16. Did the *discussion* section adequately address the limitations of the study? Was the final conclusion of the paper a balanced summary of the study findings?

rameters such as FEV^1. Further evidence of the strength of an effect is provided by dose-response relations, in which the effect estimate rises over increasing categories of exposure.

When an association appears to be present, the validity of the association must be evaluated. This can be done in studies that provide adequate detail on design and results. The internal validity should be evaluated by examining for selection bias, by misclassification, and confounding. All studies suffer to some degree from problems with validity, so a judgment must be made concerning the importance of the biases. The important biases are those that could explain the findings—that is, biases large in magnitude and operating in the direction of the finding (away from the null in positive studies, toward the null in negative studies).

Finally, the consistency of the association—that is, the repeated demonstration of a particular association in different popula-

tions and by different investigators—is valuable supporting evidence that the association truly exists. Toxicology data and reasonable consistency with a postulated biologic mechanism may also assist in determining causal associations.

Results of statistical tests of significance, probability values (p-values) or confidence intervals, usually are presented along with estimates of the relative risk. These results contribute to interpretation of studies by providing a measure of stability of the associations reported. Statistical tests evaluate the probability that the observed association could have occurred by chance alone (assuming that no effect is expected *a priori*). For example, a p-value less than .05 indicates that the likelihood of observing an effect at least as large as the one actually observed is less than 5%, given that no association truly exists. Some investigators define significance as a probability value less than .01; others require only that it be less than .10. Confidence intervals provide more information than probability values alone because they provide information on the magnitude of the association as well as the stability of the estimate.

The statistical power of a study to detect a true effect depends on the background prevalence of the disease or exposure, the size of the group studied, the length of follow-up, and the level of statistical significance required. Monitoring of a small cohort for a brief period can yield a falsely negative result. For this reason, it is important, when interpreting a negative study, to examine whether the design itself precluded a positive finding. For example, a retrospective cohort study of formaldehyde exposure had only 80% power to detect a fourfold risk in nasal cancer mortality, despite having 600,000 person-years of observation (14). The power was low because nasal cancer has a very low background prevalence. Formulas for calculating the statistical power associated with a given sample size are available in standard biostatistics and epidemiology texts.

REFERENCES

1. Blot WJ, Harrington JM, Toledo A, et al. Lung cancer after employment in shipyards during World War II. N Engl J Med 1978;299:620–624.
2. Corn M, Esmen NA. Workplace exposure zones for classification of employee exposures to physical and chemical agents. Am Ind Hyg Assoc J 1979;40: 47–60.
3. Ayer HE, Dement JM, Busch KA, et al. A monumental study: reconstruction of a 1920 granite shed. Am Ind Hyg Assoc J 1973;34:206–211.
4. Hu H, Rabinowitz M, Smith D. Bone lead as a biological marker in epidemiologic studies of chronic toxicity: conceptual paradigms. Environ Health Perspect 1998;106:1–8.
5. Creech JL, Johnson MN. Angiosarcoma of liver in the manufacture of polyvinyl chloride. J Occup Med 1974;16:150–151.
6. Billmaier D, Yee HT, Allen N, et al. Peripheral neuropathy in a coated fabrics plant. J Occup Med 1974;16:668–671.
7. Speizer FE, Wegman DH, Ramirez A. Palpitation rates associated with fluorocarbon exposure in a hospital setting. N Engl J Med 1975;292:624–626.
8. Redmond CK, Wieand HS, Rockette HE, et al. Long term mortality experience of steelworkers. Department of Health and Human Services, U.S. Government Printing Office, Washington, DC, (NIOSH) Publication No. 81-120.1981.
9. Hemminki K, Mutanen P, Saloniemi I, Niemi M-L, Vainio H. Spontaneous abortions in hospital staff engaged in sterilizing instruments with chemical agents. Br Med J 1982;285:1461–1463.
10. Wegman DH, Musk AW, Main DM, et al. Accelerated loss of FEV-l in polyurethane production workers: a four-year prospective study. Am J Ind Med 1982;3:209–215.
11. Zahm SH, Weisenburger DD, Saal RC, Vaught JB, Babbitt PA, Blair A. The role of agricultural pesticide use in the development of non-Hodgkin's lymphoma in women. Arch Environ Health 1993;48:353–358.
12. Armstrong B, Theriault G, Guenel P, Deadman J, Goldberg M, Heroux P. Association between exposure to pulsed electromagnetic fields and cancer in electric utility workers in Quebec, Canada and France. Am J Epidemiol 1994;140:805–820.
13. Eisen EA, Robins JM, Greaves IA, Wegman DH. Selection effects of repeatability criteria applied to lung spirometry. Am J Epidemiol 1984;120:734–742.
14. Blair A, Stewart P, O'Berg M, et al. Mortality among industrial workers exposed to formaldehyde. J Natl Cancer Inst 1986;76:1071–1084.
15. Monson RR. Occupational epidemiology, 2nd ed. Boca Raton, FL: CRC Press, 1989.
16. Doll R, Hill AB. Mortality in relation to smoking: ten years' observations of British doctors. Br Med J 1964;1:1399.

BIBLIOGRAPHY

Ahlbom A, Norell S. Introduction to modern epidemiology, 2nd ed. Chestnut Hill, MA: Epidemiology Resources Inc, 1990.

Beaglehole R, Bonita R, Kjellstrøm T. Basic epidemiology. Geneva: World Health Organization, 1993.
Two introductory texts on the core ideas underlying epidemiologic research and useful starting points for more advanced reading. The second book is available worldwide (through WHO), and a teacher's guide can be obtained for use with the text.

Checkoway H, Pearce NE, Crawford-Brown DJ. Research methods in occupational epidemiology. New York: Oxford University Press, 1989.
Very readable full text on epidemiologic approaches specific to occupational studies. Numerous examples are provided to guide the reader in understanding both the simple and the complex issues that must be addressed.

Hernberg S. Introduction to occupational epidemiology. Chelsea, MI: Lewis Publishers, 1992.
An excellent introductory text that is well written and illustrated. Aimed at the reader new to occupational epidemiology but somewhat familiar with principles of epidemiology.

Monson RR. Occupational epidemiology, 2nd ed. Boca Raton, FL: CRC Press, 1989.
A systematic review of methods as applied specifically to occupational settings. A practical textbook for those doing occupational studies.

Olsen J, Merletti R, Snashall D, Vuylsteek K. Searching for causes of work-related diseases: an introduction to epidemiology at the work site. Oxford: Oxford Medical Publications, 1991.
A practical introduction to epidemiology for health professionals with no formal training in the discipline. It is written to assist professionals to better plan and carry out investigation of worksite health problems.

Pagano M, Gauvreau K. Principles of biostatistics. Belmont, CA: Duxbury Press, 1993.
Basic statistics text written in a reasonable fashion with a functional index. Good general reference for statistics.

Rothman KJ, Greenland S. Modern epidemiology, 2nd ed. Philadelphia: Lippincott-Raven, 1998.
Probably the best general text on epidemiologic methods designed both for the novice and the expert. Provides principles of epidemiology in substantial detail as well as the quantitative basis for the research methods. Particularly useful as a reference.

Steenland K. Case studies in occupational epidemiology. New York: Oxford University Press, 1993.
Provides the reader the opportunity to explore further many of the questions discussed in this chapter through practical and detailed presentation of case studies of various types of epidemiologic studies.

APPENDIX
Adjustment of Rates

For purposes of illustration, adjusting for differences in age is examined in detail. Table 6-3 presents a hypothetical problem involving the myocardial infarction experience in two viscose rayon factories. To compare the incidence of myocardial infarction, a summary rate is calculated for each factory. If crude rates were calculated, it would appear that workers in factory 2 have a slightly greater risk. Comparison of these rates, however, ignores the rather striking difference in age distribution of the populations in the two factories. These can be taken into account by adjusting for age differences by either the direct method or the indirect method.

Direct Adjustment

The principle of direct adjustment is to apply the age-specific rates determined in the study groups to a set of common age weights, such as a standard age distribution. The selection of the standard is somewhat arbitrary, but often the sum of the specific age groups for the study groups is chosen. In Table 6-3, the standard population is 1,200 persons younger than 45 years and 800 persons 45 years or older. The specific rates are applied to this set of weights and then added to create an adjusted rate.

$$\text{Factory 1} = \frac{(.010 \times 1{,}200) + (.030 \times 800)}{2{,}000}$$

$$= .018$$

$$\text{Factory 2} = \frac{(.0125 \times 1{,}200) + (.050 \times 800)}{2{,}000}$$

$$= .0275$$

Not only is the magnitude of the rate of myocardial infarction affected by the adjustment procedure, but the rank order is reversed. Note that if another age distribution had been selected as the standard, the standardized rates would change. For example, for 1,500 persons younger than 45 years and 500 age 45 or older, the rate for factory 1 would become .015 and that for factory 2 would become .022. Although the absolute

magnitudes of the two adjusted rates have no inherent meaning, the relative magnitudes do. While the size of the ratio will change slightly, it will be closely duplicated regardless of the weights. In these two examples of weighting, the ratios of the adjusted rates are 1.53 and 1.47.

INDIRECT ADJUSTMENT

In indirect adjustment, standard rates are applied to the observed weights or the distribution of specific characteristics (e.g., age, sex or race) in the study populations. This provides a value for the number of cases (events) that would be expected if the standard rates were operating. The expected number of cases can be compared with the number actually observed for each study group in the form of a ratio. In Table 6-3, assume a national standard rate for myocardial infarction of 1 in 1,000 (.001) for those younger than 45 years of age and 2 in 1,000 (.002) for those 45 years or older. The expected number of cases in the two factories would then be as follows:

$$\text{Factory 1} = (.001 \times 400) + (.002 \times 600)$$
$$= 1.6$$

$$\text{Factory 2} = (.001 \times 800) + (.002 \times 200)$$
$$= 1.2$$

These expected values are compared with the observed values to calculate a standardized morbidity ratio, as follows:

$$\text{Factory 1 SMR} = \frac{22}{1.6} = 13.8$$

$$\text{Factory 2 SMR} = \frac{20}{1.2} = 16.7$$

It is tempting to compare the two SMRs and calculate a ratio similar to that calculated for the directly standardized rates. However, a drawback of indirect standardization is that SMRs cannot be compared. Because the age distributions and age-specific rates are significantly different for the two factories, the resulting comparison of the two SMRs would not distinguish differences caused by a different disease incidence rate from differences caused by a different age distribution.

It is reasonable, then, to ask why indirectly standardized rates are used. One reason is that often only one population is being studied, so comparison with the general population experience is convenient and possibly the only reasonable comparison available. Probably of greater importance is the instability of observed rates. In the example presented here, if five rather than two age groups were used and it was also necessary to adjust for both race and sex, then the total number of subdivisions necessary would be $5 \times 2 \times 2 = 20$. With a maximum of 22 cases in either factory, several of the subdivisions would contain no cases and therefore have no reliable rate estimate. Even in the illustration provided, one case more or one case less among the group of younger workers in factory 1 would have changed the age-specific incidence rate to 12.5 or 7.5, respectively, a very large difference.