Table 10-4
Selection bias in an epidemiologic study

Part 1. Unbiased Data

	Exposed	Unexposed	
Cases	20	80	Odds Ratio = (20/80) (22/178)
			OR = 2.0
Controls	22	178	

Part 2. Selection Bias: Participation is 100% for exposed cases, 80% for unexposed cases, and 80% for both exposed and unexposed controls

	Exposed	Unexposed	
Cases	20	64	Odds Ratio = (20/64) (18/142)
			OR = 2.5
Controls	18	142	

of errors in classifying subjects (or person-time) with regard to exposure, or with regard to disease. The effect of misclassification depends on whether errors in classification are differential or nondifferential. Differential errors occur if the extent of misclassification of subjects according to disease status varies by exposure category, or if the extent of misclassification of subjects by exposure status differs by disease category (diseased, not diseased). Error in disease classification is referred to as nondifferential when the extent of error is equal for exposed and unexposed subjects, and error in exposure classification is nondifferential when the extent of error in exposure classification is equal for diseased and nondiseased subject.

It is virtually certain that misclassification will occur in epidemiologic studies. This is particularly true of classification of subjects with regard to exposures that occurred in the past (i.e., before the investigators begin the study) and that depend on subjects' memories (e.g., self-reported history of smoking, of occupational exposures, of nonprescription drug use, of diet, etc.) or that must be inferred from indirect information (e.g., occupational exposure to butadiene inferred from information on particular jobs held by a subject in the synthetic rubber industry). Misclassification with respect to disease status also occurs in epidemiologic research.

Disease information may come from death certificates, which do not always accurately record decedents' medical conditions, or from clinical examinations or subjects' self reports, both of which may be subject to error.

Nondifferential misclassification errors generally are considered to have less grave consequences than differential errors (Rothman and Greenland, 1998, pp. 126-132). This is because nondifferential misclassification of a binary exposure or disease variable either produces no bias, or produces bias in a predictable direction; that is, towards the null (Birkett, 1992). For example, an IRR for a causal association, when measured in a study with nondifferential misclassification of subjects relative to a binary exposure variable (exposed v not exposed), will appear to be closer to the null value of 1.0 than it truly is. Differential misclassification, on the other hand, can either exaggerate or reduce the magnitude of an association. To minimize the potential for differential misclassification, investigators typically use objective procedures to determine subjects' exposure and disease status. That is, instead of obtaining information from subjects' unsubstantiated reports of exposure or disease, investigators preferably would use industrial hygiene measurements, laboratory tests, historical exposure or disease records, and would review and process all such data in a "blind" fashion (see later discussion).

An objective approach to exposure and disease ascertainment is, of course, highly desirable. However, objectivity merely ensures that classification errors will **tend** to be nondifferential. In small studies, it is reasonable to assume that misclassification of just a few subjects could introduce either differential or nondifferential error.

Table 10-5 illustrates nondifferential misclassification in a study examining the relationship between a binary exposure variable and mortality from a particular cause of death. Part one of the example shows the correctly classified exposure distribution of decedents and person-years. The unbiased MRR is 2.0. In part two of the example, 5% of exposed decedents and person-years and 10% of unexposed decedents and person-years are incorrectly classified with respect to exposure. The misclassification is dependent on exposure status (i.e., it is greater for the unexposed than for the exposed), but not on disease status (i.e., within each of the two exposure categories, the per cent misclassified is the same for decedents and person-years). The MRR of 1.8 is biased towards the null; that is, the MRR is closer to the null value of 1.0 than the true MRR of 2.0. A similar result would be obtained if 10% rather than 5% of exposed decedents and person-years, and 5% rather than 10% of unexposed decedents and person-years, were incorrectly classified with respect to exposure.

Table 10-6 illustrates differential misclassification in a study comparing the odds of exposure to a particular agent in a group of cases of a given disease with the odds of exposure in a group of controls without the disease of interest. Part one of the example shows the correctly classified data with an unbiased OR of 2.0. In part two, it is assumed that subjects were asked about their exposure history, and that recall of exposure was perfectly correct for 100% of both exposed and unexposed cases and for 100% of unexposed controls. However, only 82% of truly exposed controls remembered their exposure. Thus, subjects' exposure misclassification errors were differential; that is, dependent both on exposure status (more accurate for unex-

Table 10-5
Nondifferential exposure misclassification in an epidemiologic study

Part 1. Correctly classified exposure data for decedents and person-years

	Exposed	Unexposed	
Deaths	200	100	Mortality rate ratio = 200/100 **MRR = 2.0**
Person-years	100,000	100,000	
Mortality rate (MR) (x 10^5 years)	200	100	

Part 2. Biased data due to nondifferential misclassification of exposure: 5% of exposed decedents and person-years are misclassified (counted as unexposed); 10% of unexposed decedents and person-years misclassified (counted as exposed)

	Exposed	Unexposed	
Deaths	200*	100‡	Mortality rate ratio = 190/105 **MRR =1.8**
Person-years	105,000†	95,000§	
Mortality rate (MR) (x 10^5 years)	190	105	

* 200 – (0.05x200) + (0.10x100).
† 100,000 – (0.05x100,000) + (0.10x100,000).
‡ 100 - (0.10x100) + (0.05x200).
§ 100,000 – (0.10x100,000) + (0.05x100,000).

Table 10-6
Differential exposure misclassification in an epidemiologic study

Part 1. Correctly classified exposure data for cases and controls

	Exposed	Unexposed		
Cases	20	80	Odds ratio = (20/80)	(22/178)
			OR = 2.0	
Controls	22	178		

Part 2. Biased data due to differential misclassification of exposure: 100% of both exposed and unexposed cases and 100% of unexposed controls are correctly classified; 18% (4) of exposed controls are misclassified as unexposed

	Exposed	Unexposed		
Cases	20	80	Odds ratio = (20/80)	(18/182)
			OR = 2.5	
Controls	18	182		

posed) and on disease status (more accurate for cases than for controls). The biased OR in part two of the example is 2.5, and the bias is away from the null.

Many strategies of study design and data collection and development are used for the purpose of ensuring that classification errors are nondifferential. These include using comparable data collection and development techniques for the two or more groups to be compared in a study; using "blind" procedures for determining exposure and diseases status (i.e., persons responsible for classifying subjects according to exposure are unaware of subjects' disease status, and persons responsible for classifying subjects according to disease status are unaware of subjects' exposure status); and using objective records, rather than self-reports, to determine subjects' exposure and disease status.

Another approach that is controversial but continues to be used involves selecting study groups that are likely to be comparable in terms of the extent of misclassification in their data (Greenland and Robins, 1985). For example, if all cases in a case-control study are deceased, an investigator may choose to require that all controls also be deceased. The rationale is that the two decedent groups will have a similar degree of misclassification in data, such as information on smoking and lifetime job history, that must be furnished by surrogates. If some cases are

alive at the time of data collection and some are deceased, an investigator may match cases and controls on vital status, choosing living controls for cases who are alive and deceased controls for cases who are dead.

Confounding

Confounding refers to an error that occurs when the groups being compared in a study differ with respect to disease determinants other than the factor whose possible effect is being assessed. When an association is confounded, the estimated effect of the exposure is distorted because it is mixed with the effect of another factor. In order to be a confounder a factor must be predictive of the occurrence of the disease under study and must be associated with the exposure under study, as explained further in the next example. The impact of confounding can be large or small and can produce an under- or over-estimate of the effect of the exposure of interest.

Table 10-7 illustrates confounding by age in a hypothetical study of the relationship between occupational exposure to bis-chloromethyl ether and the occurrence of lung cancer. In this example, the rate of lung cancer is three times higher among the exposed than among the unexposed **within each age group** (i.e., each of the three age-specific IRRs is 3.0). However, if one ignores age and computes the rate for the entire exposed group (i.e., for the "total") as done in the first part of this example, this rate is 51 cases/64,000 person-years, for an IR of 79.7/100,000 years. Again, if one ignores age and computes the rate for the entire unexposed group, the IR is 13 cases/35,000 person-years, or 37.1/100,000 years. The IRR is 79.7/37.1 or 2.1, indicating erroneously that the rate of lung cancer is only two times higher among the exposed than among the unexposed. In this example, age is a confounder:

- age is "predictive of" the lung cancer IR – the older the age, the higher the IR; and
- age is associated with exposure – the exposed group is younger than the unexposed group, since 70% of the person-years of the exposed are in the youngest age group, whereas only 57% of the person-years of the unexposed are in the youngest age group.

Epidemiologists use several strategies to control for potential confounding. These consist of restricting a study to subjects in one category of a potential confounder (e.g., if a study is restricted to men; gender cannot be a confounder) and adjusting measures of association for potential confounding by using certain analytic techniques, such as standardization, other stratified analysis procedures, or multivariable modeling. The main purpose of these analytic procedures is to remove distortion due to the groups' being compared having different distributions of causal factors other than the particular factor under investigation.

A measure of occurrence or of association that does not take other factors (potential confounders) into consideration is referred to as "crude." For example, the data in Table 10-7 indicate that the crude IRR (i.e., the IRR computed ignoring age) is 2.1. A measure of occurrence or of association that uses stratified or multivariable analysis to consider a factor or fac-

Table 10-7
Confounding by age in a study of bis-chloromethyl ether (BCME) exposure and lung cancer

Age group	Exposed to asbestos: Obs.* cases	Exposed to asbestos: Person-years	Exposed to asbestos: IR (incidence rate)	Unexposed to asbestos: Obs. cases	Unexposed to asbestos: Person-years	Unexposed to asbestos: IR (incidence rate)	Age-specific IRR (rate ratio)
20-39 yrs	24	40,000	$60/10^5$ y	400	2,000,000	$20/10^5$ y	3.0
40-59 yrs	18	15,000	$120/10^5$ y	400	1,000,000	$40/10^5$ y	3.0
60-89 yrs	9	3,000	$300/10^5$ y	500	500,000	$100/10^5$	3.0
Total	51	64,000	79.7	1,300	3,500,000	37.1	**2.1 (crude IRR)**

Standardization of the IRR to obtain a standardized incidence ratio (SIR)

Age group	Obs.* cases	Expected number (in each age group, the person-years of the exposed, multiplied by the rate of the unexposed†)	IRR (Obs/Exp no.)
20-39 yrs	24	40,000 PY x $20/10^5$ y = 8	24/8 = 3.0
40-59 yrs	18	15,000 PY x $40/10^5$ y = 6	18/6 = 3.0
60-89 yrs	9	3,000 PY x $100/10^5$ y = 3	9/3 = 3.0
Total	51	8 + 6 + 3 = 17	**51/17 = 3.0 (age-adjusted IRR= SIR)**

* Obs, observed number of cases of lung cancer.
† Expected, expected number of cases of lung cancer in the exposed, if the exposed had the same lung cancer rate and same age distribution as the unexposed.

tors other than the particular one under investigation is referred to as "adjusted" for the other factor(s).

As shown in part two of the example in Table 10-7, an adjusted IRR would be 3.0, obtained by dividing the observed number of lung cancer cases among the exposed by the expected number, 51/17 = 3.0. Table 10-7 illustrates the computation of an age-adjusted IRR for the exposed compared to the unexposed, obtained using a technique called standardization. The latter technique is a form of stratified analysis. In standardization, one computes an "expected" number of cases of lung cancer for each age group of the exposed. The expected number is the number of cases that would occur in the exposed if they retained their own age distribution but had the same age-specific lung cancer IRs as the unexposed group. Such an IRR is referred to as a standardized IR (SIR) or standardized MR (SMR). In this example, the SIR is 3.0, exactly as one would expect, given that each of the three age-specific IRRs is 3.0. This SIR also can be viewed as a weighted average of the age-specific IRRs, with the weight for each age-specific IRR equal to the person-years of the exposed in the corresponding age group, multiplied by the age-specific rate in the unexposed (see Table 10-7, part 2).

Confounding can be either negative or positive (Rosner, 1995, p. 401). Negative confounding, when the crude is lower than the adjusted measure of association, occurs if the confounder is associated positively with disease and negatively with exposure, or if the confounder is associated negatively with disease and positively with exposure. The above example illustrates negative confounding. The crude IRR of 2.1 is lower than the adjusted IRR of 3.0, and the confounder, age, is associated positively with lung cancer and negatively with exposure. Positive confounding, when the crude is higher than the adjusted measure of association, occurs either if the confounder is positively associated both with disease and with exposure, or if the confounder is negatively associated both with disease and with exposure.

PRECISION

Precision refers to lack of random error in the measurement of an association, and random error refers to that part of an experience that is due to chance (see Volume 1, Chapter 7). Random error can lead to the appearance of an effect, even when the exposure under investigation actually does not cause the disease of interest. In interpreting any association observed in a particular epidemiologic study, it is important to consider, among other things, how likely it is that the association could have occurred by chance, if no causal effect exists. Similarly, when no association is seen, it is important to consider whether a true effect could have been missed simply by chance; that is, simply because the study was too small. Measurement of the precision of IRRs, MRRs, IPRs or rate or risk differences allows one to evaluate the role of chance in producing an apparent relationship or in failing to find a true effect.

Precision depends on the size of the study, on the relative size of the groups being compared, and on the frequency of exposure and of disease in the populations studied. Precision can be improved by increasing the number of subjects or other units of observation to be in-

cluded in an investigation (Rothman and Greenland, 1998, pp. 135-137). When cost and effort constraints make it infeasible to include larger numbers of subjects, precision can be enhanced by focusing on a study base in which the disease and/or the exposure of interest are relatively common. For example, one might choose to restrict a study according to age, including only older persons in whom disease rates tend to be relatively high; or one might choose to restrict a study according to geographic region, including only those persons living in a region where the exposure of interest is relatively common. Matching (defined briefly in a later section) groups to be compared in a study on the basis of potential confounders may also be used to increase precision.

The precision of estimates of IRRs, MRRs, or IPRs, of rate or risk differences, of ORs, and of other measures of association of interest in epidemiology typically is expressed as a confidence interval or as a standard error. P-values also may be used. Confidence intervals and p-values are discussed briefly below. Computation of these measures is beyond the scope of this chapter, but details are provided in epidemiology text books (Checkoway et al., 1989; Rothman, 1986; Rothman and Greenland, 1998).

Confidence Interval

A confidence interval (CI) specifies the range of values of a true effect with which the data from a particular study are compatible. The width of the CI of a particular measure of association is determined by the extent of random variability in the study and by an arbitrary "confidence level," typically 95% in observational epidemiologic research. For example, in a study of synthetic rubber industry workers, the MRR for leukemia among hourly workers compared to the general population was 1.4, with a 95% CI of 1.0-1.9. Assuming the point estimate of 1.4 is valid, this CI indicates that it is 95% likely that the true value of the MRR lies in the range of 1.0 to 1.9. If a series of similar investigations of synthetic rubber industry workers were conducted, the MRR would be in the 1.0-1.9 range in 95% of the studies.

P-Value

Another measure of random error is the P-value, computed from a statistical hypothesis test. P often is construed as the probability that a test statistic as large or larger than that found in a study would arise by chance alone if the exposure and disease were not, in fact, associated. P-values range from 0 to 1, with small P-values often taken as indicating that the observed results have little compatibility with the null hypothesis. Usually, a P-value of 0.05 or lower is considered statistically significant, implying that the result probably did not occur by chance. These definitions and interpretations of P-values are conventional, but inaccurate, and can be misleading (Rothman and Greenland, 1998, pp. 185-186). CIs and standard errors are considered to be preferable to P-values in expressing the amount of precision in epidemiologic studies (Lang et al., 1998).

Power

Power indicates the adequacy of a given study size (number of subjects or amount of person-time) for detecting an exposure-disease association of a certain minimum magnitude, if in fact a true effect exists. Consideration of power can be useful for research planning purposes. This is because a power calculation indicates whether a study of a particular size will yield statistically significant results, given an outcome of interest of a particular frequency, and given an assumption about the magnitude of the association to be investigated and about the statistical tests to be performed. If, for example, exposure to ethylene glycol ether is assumed to produce a fivefold increase in the IP for congenital malformation, a smaller number of pregnancies can be studied than if ethylene glycol ether is assumed to produce a twofold increase in IP. Similarly, a statistically informative study of a medical condition occurring in occupational groups will require fewer subjects if the condition is common than if the condition is rare. After a study has been conducted, the issue of power has little relevance; instead, the best way to assess precision is to examine CIs (Checkoway et al., 1989, pp. 76-77).

MAJOR EPIDEMIOLOGIC STUDY DESIGNS

Most research in occupational epidemiology is nonexperimental. Because investigators do not manipulate the exposure conditions experienced by study subjects, nonexperimental research is more prone to systematic error and to confounding than are experiments. Each of the two main study designs used in epidemiology (the follow-up and the case-control study designs) has specific uses and areas of application, specific measures of association, and particular advantages and disadvantages. Also, both follow-up and case-control studies have several important variants.

Follow-up Studies

In a follow-up study, a group of individuals who have a characteristic in common (the "study group" or "subjects", sometimes referred to as a "cohort") and who are initially free of the disease of interest is identified. Each subject is monitored ("followed up") over time to determine changes in exposure status. This information may be used to form exposed and unexposed subgroups of subjects or subgroups classified according to level of exposure; alternatively, all members may be considered "exposed." Each person in the study group is followed up over time to determine whether he or she develops the disease of interest, information that is used to measure the groups' morbidity and/or mortality rates. The subsequent analysis determines if the IR, MR, or IP of exposed subjects is higher, lower or the same as that of unexposed subjects.

A follow-up study may involve open or closed populations (subjects). In studies of open groups, the unit of observation usually is a unit of person-time, such as a person-year, and the measure of association is an IRR or an MRR. In follow-up studies of closed groups, the unit of observation may be either person-time or the individual subject, and correspondingly, the mea-

sure of association is an IRR or MRR or an IPR. Most follow-up studies in the occupational setting are of open study groups, and most estimate IRRs or MRRs as measures of association.

A follow-up study may be retrospective (historical) or prospective. In retrospective follow-up studies, subjects are identified as of some time in the past and are followed into the present, often over several decades. In prospective follow-up studies, the study group is enumerated currently and followed into the future. Prospective follow-up studies investigating the relationship between exposures and chronic diseases may continue for many decades. Some studies have both retrospective and prospective components.

Retrospective follow-up studies of completely closed study groups are rare in occupational epidemiology, whereas prospective follow-up studies of closed study groups are more common. In occupational epidemiology, follow-up studies are used to investigate the causes of cancer, chronic respiratory disease, and many other chronic diseases. Retrospective follow-up studies have been used particularly often to investigate fatal diseases, because mortality registries, which facilitate such studies, have existed in many countries for many decades.

Retrospective follow-up studies have not been as useful for determining the etiology of nonfatal diseases or physiologic dysfunction (e.g., pulmonary, renal, neurologic dysfunction). These are best suited to prospective follow-up studies. Chronic diseases with long induction times (amount of time required from first exposure to a causal agent to the development or detection of overt, irreversible disease) have most often been studied with retrospective follow-up studies. This is because prospective follow-up study design would require enormous resources to observe large numbers of subjects into the future to determine their exposures and disease experience, with evaluation of hypotheses possible only after much elapsed time.

The key steps required to conduct follow-up studies are:

- Specification of objectives
- Identification of study group members and specification of exposed groups
- Identification of comparison groups
- Follow-up or tracing of subjects to determine their vital status, their disease status and their individual person-years of observation
- Analysis, or comparison of disease rates between exposed and unexposed groups or across groups specified on the basis of level of exposure; control of confounding; examination of the cause-to-effect time sequence (induction time effects)
- Interpretation of results.

Follow-up studies may have a rather broad, descriptive purpose, such as determining if workers in a particular plant, group of plants, or occupation have any unusual disease occurrences. Alternatively, follow-up studies may be done to evaluate a specific hypothesis, such as the hypothesis that exposure to butadiene causes leukemia in humans (Delzell et al., 1996; Macaluso et al., 1996). A clear statement of objectives and rationale helps to guide design, data collection, analysis and interpretative decisions (Epidemiology Task Group, 1991).