

Dealing with Biases

Two broad types of error afflict epidemiologic studies: random error and systematic error. In designing a study, an epidemiologist attempts to reduce both sources of error. In interpreting a study, a reader should be aware of both types of error and how they have been addressed.

What is meant by error in a study? In Chapter 5, I said that an epidemiologic study could be viewed as an attempt to obtain an epidemiologic measure. The object of measurement may be a rate or a risk, but it typically is a measure of effect, such as an incidence rate ratio. Suppose a study is conducted to attempt to measure the ratio of the incidence rate of Alzheimer's disease among those who are physically active compared with those who are physically inactive. We can imagine that there is a correct value for the incidence rate ratio. A given study will produce an estimate of this correct value. If the study estimates a value of the incidence rate ratio that is close to the correct value, we would consider the study to be accurate, which means that it has little error. Conversely, a study estimate that differs considerably from the correct value is inaccurate. Unfortunately, we can never know the correct value for the rate ratio of Alzheimer's disease among physically active people compared with the physically inactive or for any other measure that we try to estimate; all we can know is the value of the estimates from a study. Because the correct values are unknown, we cannot determine the actual amount of error in any given study. Nevertheless, epidemiologists can still take steps in the design and analysis of studies to reduce errors. We also can look for features in the design and analysis of a study that may contribute to or prevent errors.

This chapter focuses on systematic error; random error is discussed in Chapter 8, which deals with statistical issues in epidemiologic research. Another term for systematic error is *bias*. Bias can refer to an attitude on the part of the investigator, but it is also used to describe any systematic error in a study. A study can be biased because of the way in which the subjects have been selected, the way the study variables are measured, or some confounding factor that is not completely controlled.

There is a simple way to distinguish random errors from systematic errors. Imagine that a given study could be increased in size until it was infinitely large.

Some errors would be reduced to zero if a study became infinitely large; these are the random errors. Other errors are not affected by increasing the size of the study. Errors that remain even in an infinitely large study are the systematic errors, also described as biases (Fig. 7-1). As study size increases and as random error concomitantly decreases, the relative role of systematic error becomes greater. In a sufficiently large study, virtually all errors of concern are systematic errors.

To see the difference between systematic errors and random errors, consider the following example. Suppose your task is to determine the average height of women in the city of Centerville, which has a population of 500,000 women. To conduct this work, you are supplied with an official measuring tape. You may decide to measure the height of 100 women sampled randomly from the population of all women in the city. You can use the average of the 100 measurements as an estimate of the average height of women in Centerville. What sources of error affect your estimate? A measuring tape will give different readings depending on how it is held, how it is read, the time of day the measurement is taken, and who is taking the measurement. Some of these errors, such as how the measuring tape is held during a given measurement, may be random; some of these errors sometimes lead to a reading that is too high and sometimes to a reading that is too low, but on average, readings do not tend to be too high or too low. If the sample of 100 were increased to 1000 or to 10,000 women, the effect of these random errors would become less important because the greater number of measurements would ensure that the discrepancy between the average measured height for women in the sample and the height of all women in Centerville would be close to zero. Other errors, however, would not be affected by increasing the number of women measured. Suppose that the official tape used in the measurements was a cloth tape that had been laundered before the project began. Unknown to anyone, the laundering shrank the tape. Consequently, the height

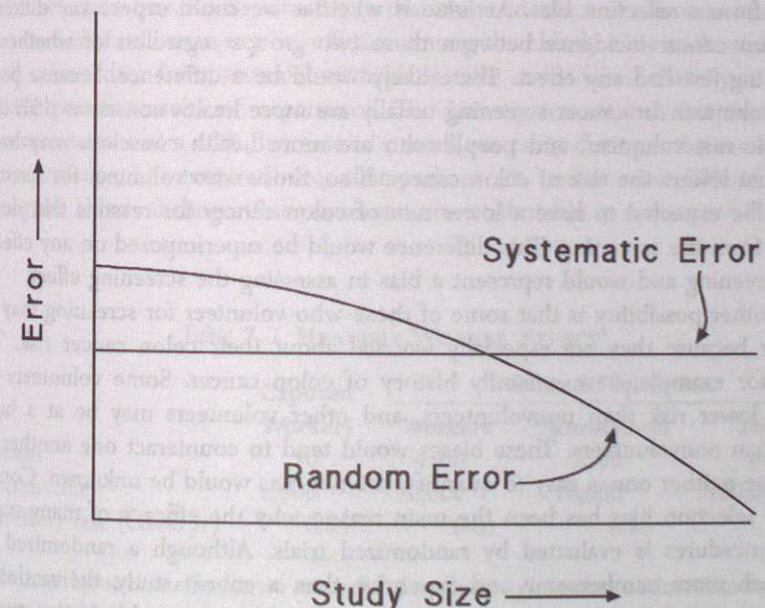


Figure 7-1 The relation of systematic error and random error to study size.

Dealing with Biases

Two broad types of error afflict epidemiologic studies: random error and systematic error. In designing a study, an epidemiologist attempts to reduce both sources of error. In interpreting a study, a reader should be aware of both types of error and how they have been addressed.

What is meant by error in a study? In Chapter 5, I said that an epidemiologic study could be viewed as an attempt to obtain an epidemiologic measure. The object of measurement may be a rate or a risk, but it typically is a measure of effect, such as an incidence rate ratio. Suppose a study is conducted to attempt to measure the ratio of the incidence rate of Alzheimer's disease among those who are physically active compared with those who are physically inactive. We can imagine that there is a correct value for the incidence rate ratio. A given study will produce an estimate of this correct value. If the study estimates a value of the incidence rate ratio that is close to the correct value, we would consider the study to be accurate, which means that it has little error. Conversely, a study estimate that differs considerably from the correct value is inaccurate. Unfortunately, we can never know the correct value for the rate ratio of Alzheimer's disease among physically active people compared with the physically inactive or for any other measure that we try to estimate; all we can know is the value of the estimates from a study. Because the correct values are unknown, we cannot determine the actual amount of error in any given study. Nevertheless, epidemiologists can still take steps in the design and analysis of studies to reduce errors. We also can look for features in the design and analysis of a study that may contribute to or prevent errors.

This chapter focuses on systematic error; random error is discussed in Chapter 8, which deals with statistical issues in epidemiologic research. Another term for systematic error is *bias*. Bias can refer to an attitude on the part of the investigator, but it is also used to describe any systematic error in a study. A study can be biased because of the way in which the subjects have been selected, the way the study variables are measured, or some confounding factor that is not completely controlled.

There is a simple way to distinguish random errors from systematic errors. Imagine that a given study could be increased in size until it was infinitely large.

Some errors would be reduced to zero if a study became infinitely large; these are the random errors. Other errors are not affected by increasing the size of the study. Errors that remain even in an infinitely large study are the systematic errors, also described as biases (Fig. 7-1). As study size increases and as random error concomitantly decreases, the relative role of systematic error becomes greater. In a sufficiently large study, virtually all errors of concern are systematic errors.

To see the difference between systematic errors and random errors, consider the following example. Suppose your task is to determine the average height of women in the city of Centerville, which has a population of 500,000 women. To conduct this work, you are supplied with an official measuring tape. You may decide to measure the height of 100 women sampled randomly from the population of all women in the city. You can use the average of the 100 measurements as an estimate of the average height of women in Centerville. What sources of error affect your estimate? A measuring tape will give different readings depending on how it is held, how it is read, the time of day the measurement is taken, and who is taking the measurement. Some of these errors, such as how the measuring tape is held during a given measurement, may be random; some of these errors sometimes lead to a reading that is too high and sometimes to a reading that is too low, but on average, readings do not tend to be too high or too low. If the sample of 100 were increased to 1000 or to 10,000 women, the effect of these random errors would become less important because the greater number of measurements would ensure that the discrepancy between the average measured height for women in the sample and the height of all women in Centerville would be close to zero. Other errors, however, would not be affected by increasing the number of women measured. Suppose that the official tape used in the measurements was a cloth tape that had been laundered before the project began. Unknown to anyone, the laundering shrank the tape. Consequently, the height

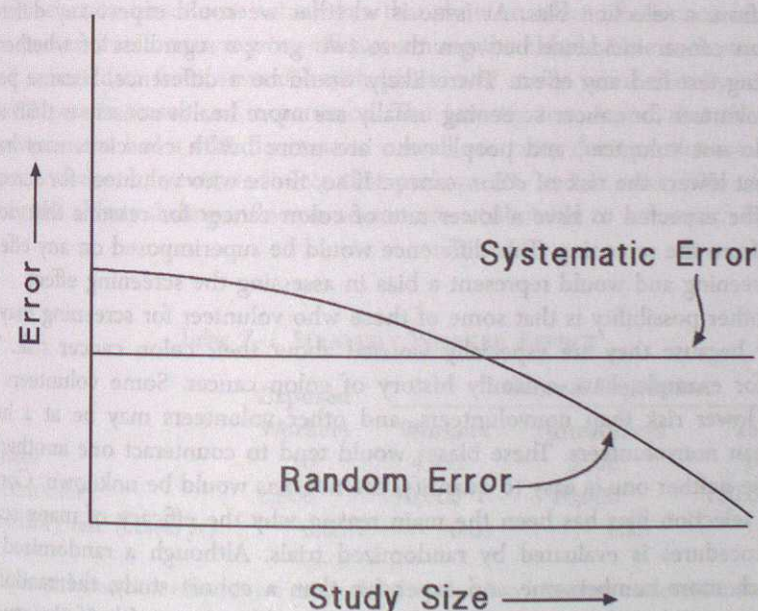


Figure 7-1 The relation of systematic error and random error to study size.

estimates derived from using the shrunken tape would tend to be high by an amount that depends on the amount of shrinkage. This systematic error cannot be reduced by taking more measurements with the same shrunken tape. Similarly, any defect in the measuring technique, such as a tendency to hold the tape crookedly, may also lead to measurements that are systematically wrong and would not be offset by increasing the number of subjects.

SOURCES OF BIAS IN EPIDEMIOLOGIC STUDIES

Error can creep into epidemiologic studies from myriad directions. Although many types of specific biases have been described, it is helpful to classify bias into three broad categories: selection bias, information bias, and confounding.

Selection Bias

Selection bias is a systematic error in a study that stems from the procedures used to select subjects and from factors that influence study participation. It comes about when the association between exposure and disease differs for those who participate and those who do not participate in the study. Because the association between exposure and disease among nonparticipants is unknown, the presence of selection bias must usually be inferred, rather than observed.

Suppose that a new screening test was devised to detect colon cancer and that this test was offered to a community in a pilot evaluation. Later, the efficacy of the test was assessed by comparing the incidence rate of colon cancer among those who volunteered to be tested with the incidence rate among community residents who were not tested. We would suspect that such a comparison would suffer from a selection bias. At issue is whether we could expect any difference in colon cancer incidence between these two groups regardless of whether the screening test had any effect. There likely would be a difference, because people who volunteer for cancer screening usually are more health conscious than those who do not volunteer, and people who are more health conscious may have a diet that lowers the risk of colon cancer. If so, those who volunteer for screening might be expected to have a lower rate of colon cancer for reasons that do not result from the screening. This difference would be superimposed on any effect of the screening and would represent a bias in assessing the screening effect.

Another possibility is that some of those who volunteer for screening may volunteer because they are especially worried about their colon cancer risk. They may, for example, have a family history of colon cancer. Some volunteers may be at lower risk than nonvolunteers, and other volunteers may be at a higher risk than nonvolunteers. These biases would tend to counteract one another, but because neither one is easy to quantify, the net bias would be unknown. Concern about selection bias has been the main reason why the efficacy of many screening procedures is evaluated by randomized trials. Although a randomized trial is much more cumbersome and expensive than a cohort study, the randomization ensures that the groups studied are reasonably comparable if the study is reasonably large.

The selection bias in the previous example is a bias arising from self-selection, because the study subjects selected themselves to be screened. Selection bias can also arise from choices made more directly by the investigator. For example, many studies of workers' health have compared the death rate among workers in a specific job with that among the general population. This comparison is biased because the general population contains many people who cannot work because of ill health. Consequently, overall death rates for workers are often substantially lower than death rates for the general population, and any direct comparison of the two groups is biased. This selection bias is often referred to as the *healthy worker effect*. One way to avert the bias is to compare the workers in a specific job with workers in other jobs that differ in their occupational exposures or hazards. If all subjects involved in the comparison are workers, the investigator can avoid bias from the healthy worker effect.

Table 7-1 shows how the healthy worker effect comes about. If the mortality rate of an exposed group of workers at a specific plant is compared with that of the general population (the Total column in Table 7-1), their overall mortality rate appears much lower; in this hypothetical example, their overall mortality rate is 5/7, or 71% of the rate in the general population. The general population, however, comprises two groups: a majority that is healthy enough to work and a minority that is too ill to work. The latter group is included among the nonworkers in Table 7-1, and results in the nonworkers having a higher mortality rate than the remainder of the general population that comprises current workers. In this hypothetical example, workers in the general population have the same mortality rate as the exposed workers at the study plant, but because the nonworkers in the general population have a rate that is five times as great as that of workers, the overall rate in the general population is considerably greater than that of the exposed workers. In a study that compared the mortality rate of the exposed workers with that of the general population, the exposed workers would have a lower mortality rate as a result of this selection bias.

The data in Table 7-1 are hypothetical data chosen to illustrate the healthy worker selection bias. Some actual data that show an effect of selection bias come from studies of influenza vaccine efficacy among the elderly. One such study combined cohort data from several health plans over many years, thereby including 713,872 person-seasons of experience. The investigators found that those who were vaccinated had a 48% decrease in overall mortality during influenza season.¹ Other evidence, however, indicates that perhaps 5% and at most 10% of deaths among the elderly during influenza season are attributable to influenza.² How can

Table 7-1 HEALTHY WORKER EFFECT^a

	Exposed Workers	General Population		
		Workers	Nonworkers	Total
Deaths	50	4500	2500	7000
Person-time	1,000	90,000	10,000	100,000
Mortality rate (cases/yr)	0.05	0.05	0.25	0.07

^aThe healthy worker effect is an example of a selection bias that underestimates the mortality related to occupational exposures, as illustrated by these hypothetical rates for workers and the general population.

a vaccine, even one that is completely effective in preventing influenza, prevent one half of all deaths among those vaccinated if influenza itself accounts for at most 10% of those deaths? In simple terms, it cannot. The huge decrease in overall mortality must reflect a selection bias. The unvaccinated group in this study is likely to include most of the elderly patients who are at the brink of death, because for them vaccination is most likely not seen as a priority. Evidence to support this hypothesis about selection bias comes from a study that examined the effect of influenza vaccine on mortality in the elderly during three periods each year: before influenza arrived, during influenza season, and after the influenza season.³ The findings are summarized in Figure 7-2. During the influenza period, the investigators found about the same vaccine efficacy as had been reported by others, with almost 50% of all deaths being "prevented." They also found, however, that the effect estimate indicating protection from the vaccine for all causes of death appeared even greater during the period each winter before the influenza virus arrived, when the vaccine could not have had any effect and all the apparent effect must stem from bias. Smaller biases were evident after the influenza season, when the vaccine efficacy should also theoretically be zero. The biases observed before and during the influenza season would also be operating during the influenza season. The selection factors responsible for the bias appear to be strongest at the outset of the study period, just after the vaccinations are administered, and they then weaken with the passage of time, as can be expected if the effect is the result of vaccine not being offered to those with a high risk of near-term death. A similar trend, also indicating strong selection bias, was evident for the outcome of hospitalization for pneumonia or influenza. These data indicate that selection bias among the elderly getting vaccinated for influenza is much stronger than any possible effect of the vaccine. These findings leave open the question of the actual magnitude of the vaccine effect on death in the elderly.

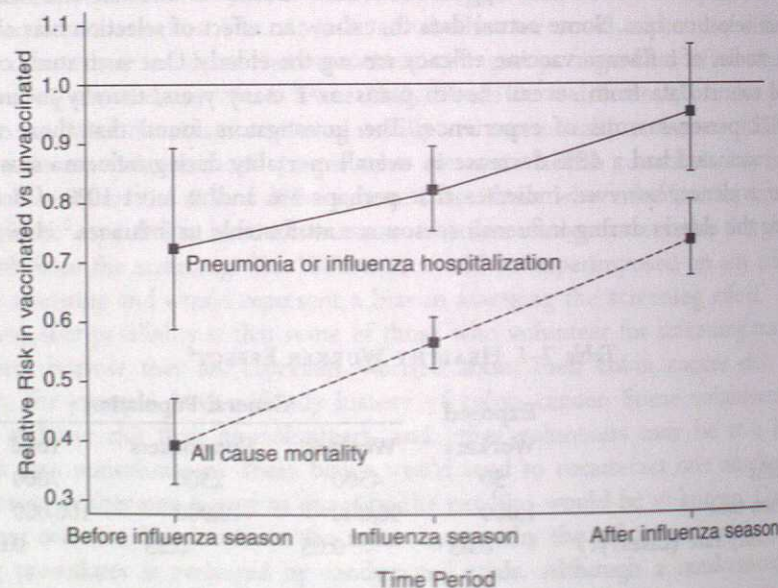


Figure 7-2 The relation of systematic error and random error to study size. (Reproduced with permission from Jackson et al.³)

MATCHING IN CASE-CONTROL STUDIES

Another prominent form of selection bias comes from a design feature of some case-control studies. When *matching* is used to select controls in case-control studies, ostensibly to prevent confounding, matching usually paradoxically results in selection bias. In Chapter 5, the point was made that to get a valid effect estimate in a case-control study, the controls must be sampled independently of the exposure. Matching in case-control studies typically violates this assumption. With matching, controls are selected because they have one or more characteristics that match the corresponding characteristics of a case in the study. Commonly used matching factors are age, sex, and geographic location, but they may also include many other factors that may be specific to a given study. The motivation for matching usually is to prevent confounding, and the matching factors therefore are usually potential confounding factors. As described later in this chapter, confounding factors are associated with both exposure and disease. By matching controls to cases on possible confounding factors, the investigator selects controls according to factors that are related to exposure, depending on the extent that the exposure is associated with the matching factors. Because of the matching, the exposure distribution in the control series may not reflect the exposure distribution of the source population for cases. Instead, the exposure distribution among matched controls will tend toward the exposure distribution of the cases. If the exposure were perfectly correlated with one of the matching factors, controls would then have exactly the same exposure distribution as the cases, which would appear to indicate no effect of exposure, regardless of the actual effect that the exposure has.

Suppose 10% of a community is exposed to an agent that multiplies the risk of disease tenfold. Let us hypothesize that males have a five times greater risk for the disease than females, and 90% of males are exposed, compared with only 10% of females. If we have 100,000 males and 100,000 females in the population, the data for this community describing the risk during 1 year is summarized in Table 7-2. Because males and exposed people are at higher risk, and most males are exposed, most cases occur in exposed males.

Because males have a much greater risk of disease than females and because the preponderance of exposed people are males, whereas most unexposed people are females, the imbalance of males between exposed and unexposed subgroups will confound the effect of exposure. Although the effect of exposure is to increase the risk of disease tenfold, if we calculated the risk among all exposed, 4600/100,000,

Table 7-2 HYPOTHETICAL DATA SHOWING RISK FOR A DISEASE DURING 1 YEAR BY EXPOSURE STATUS AND SEX

	Sex ^a	Population	Data Risk	No. of Cases
Exposed	Male	90,000	5.00%	4500
	Female	10,000	1.00%	100
Unexposed	Male	10,000	0.50%	50
	Female	90,000	0.10%	90

^aBeing male is associated with exposure and is a risk factor for disease.

and compared it with the risk among all unexposed, 140/100,000, we would obtain a risk ratio estimate of 32.9, rather than the value of 10 that corresponds to the actual effect of exposure.

Can matching prevent confounding? We could try conducting a cohort study within this population and match the unexposed subjects to the exposed subjects by sex. We could take a random 10% sample of the total exposed population as the exposed cohort and match 10,000 unexposed people to this group of 10,000 exposed people so that each person in a matched pair has the same sex. The summary data after this matching is given in Table 7-3.

After matching by sex, there is no longer an imbalance of males between exposed and unexposed. The crude data for the study represented in Table 7-3 produce an estimate of risk among exposed of 460/10,000 and among unexposed of 46/10,000, for a risk ratio of 10. Matching has prevented the confounding by male sex.

The situation is not so pretty, however, if a case-control study is conducted with the aim of using matching to prevent confounding by male sex. Suppose we include in such a study all the cases occurring during 1 year in the community. From Table 7-2, we have a total of 4740 cases. We can then select a control group of 4740 people from the community, matched by sex, in an attempt to prevent confounding. Ideally, the controls should be selected from the entire population at risk to be cases, which in this setting is the entire population of the community, rather than just the noncases (see Chapter 5). Of the 4740 cases, 4550 are males and 190 are females. The 4740 controls, after matching by sex, include 4550 male controls and 190 female controls. Because 90% of males are exposed and 10% of females are exposed, we would expect, on average, that there would be $0.9 \times 4550 = 4095$ exposed male controls and $0.1 \times 190 = 19$ exposed female controls, for a total of $4095 + 19 = 4114$ exposed controls. The summary data for this case-control study is given in Table 7-4.

Unlike the cohort study, matching in the case-control study does not give the correct risk ratio of 10. The estimate also differs from that of the confounded relation, $RR = 32.9$, which was seen for the total population. The RR estimate of 5.0 that is obtained is an underestimate of the correct value, rather than an overestimate. What happened? This result stems from selection bias. Choosing controls based on their sex leads to a control series that is mostly male, and that has an exposure distribution that has been shifted toward that of the cases, leading to an underestimate of the effect. The matching has substituted one bias for

Table 7-3 HYPOTHETICAL COHORT STUDY^a

	Sex	Population	Data Risk	No. of Cases
Exposed	Male	9,000	5.00%	450
	Female	1,000	1.00%	10
Unexposed	Male	9,000	0.50%	45
	Female	1,000	0.10%	1

^aBased on a 10% sample of exposed from the population in Table 7-2 and 10,000 unexposed people matched by sex.

Table 7-4 HYPOTHETICAL CASE-CONTROL STUDY BASED ON ALL CASES IN THE STUDY POPULATION AND ONE CONTROL PER CASE MATCHED BY SEX

	Exposed	Unexposed	Total
Cases	4,600	140	4,740
Controls	4,114	626	4,740

another. As is explained in Chapter 5, the key design element in a case-control study is that controls must be selected independently of exposure. If they are not, selection bias results. Matching controls to cases for a variable that is correlated with exposure introduces selection bias because it violates the design element that controls must be selected independently of exposure.

In Table 7-5, the case-control data are shown separately for males and females. Among males and among females, the case-control data give the correct estimate of $RR = 10$. This analysis illustrates that in case-control studies the selection bias introduced by the matching can be removed by appropriate analytic methods, such as stratifying the data by the matching factor or factors. Regression models that condition on the matching factors can also be used to remove the selection bias.

As explained in Chapter 10, stratifying the data into male and female groups can suffice to control confounding by sex, even without matching by sex in subject selection. Because matching by sex introduces a bias that also requires control of sex in the data analysis to be removed, what does matching by sex in subject selection achieve? The answer is very little and perhaps nothing. One argument for matching in a case-control study is that the data analysis becomes more efficient in a technical sense. The distribution of controls over the two strata in Table 7-5 is identical to that of the cases as a result of the matching. Having controls distributed across strata identically to cases ordinarily makes for a statistically efficient stratified analysis. Without matching, one half of controls would have been female, and there would have been more than 2000 female controls to compare with only 190 female cases, whereas male cases would have outnumbered male controls. Although matching in a case-control study does not appear to improve validity (by preventing confounding), it may improve the efficiency of a stratified analysis that is employed to remove the confounding.

Unfortunately, the argument for efficiency gain from matching in case-control studies is not clear-cut. One problem is that analytic control of the matched variable may not have even been necessary without the matching. If the matched variable is related to exposure, matching on it will introduce selection bias. But if it is not related to disease, it is not a confounding factor and can be ignored. Matching

Table 7-5 CASE-CONTROL DATA FROM TABLE 7-4, STRATIFIED BY SEX

	Males		Females	
	Exposed	Unexposed	Exposed	Unexposed
Cases	4,500	50	100	90
Controls	4,095	455	19	171
	RR = 10		RR = 10	

for such a variable introduces the need to control it in the analysis, which typically cannot improve efficiency compared with not needing to control for that factor in the first place. Another problem is that matching on some variables or a set of variables may lead to small numbers within strata. In the illustration, we matched on sex and combined all matched pairs that were male into a male stratum and all female pairs into a female stratum. If matching is implemented for many variables, however, it may produce unique combinations of values for each matched set, leading to strata for the stratified analysis that will each have a single case with one or more matched controls. With such small numbers in the strata, there is a reasonably high likelihood that the case and all the matched controls within a set will have the same value for exposure: all exposed or all unexposed. When exposure does not vary within a stratum, the stratum does not contribute information to the analysis. Effectively, any subjects in such a stratum, which is described as a *concordant set*, are lost to the analysis, leading to a loss of efficiency.

Given these potential problems, is matching worthwhile in case-control studies? Sometimes it is, but often it is not. Matching can be expensive, and in case-control studies, it does not improve validity. Efficiency gains are possible but not guaranteed and may not be worth the added cost. In some settings, efficiency may be lost rather than gained because of concordant sets. For these reasons, matching usually is best avoided in case-control studies, except for some specific exceptions. One exception is *convenience matching*. There may be circumstances in which some types of matching may simply be a convenient way to identify controls. *Risk-set sampling* is an example of a type of matching often done for convenience (see Chapter 5); it involves matching on time as a means of selecting controls proportional to their person-time contribution to the source population of cases. With convenience matching, the matching factor may not be related to exposure, and the matching may not introduce any selection bias. In that event, it can be ignored in the analysis. On the other hand, if the time variable in risk-set sampling is related to exposure, it must be controlled in the analysis, as is the case for any matching factor in a case-control study that is related to exposure. For example, consider a case-control study of mobile telephone use and brain cancer that matched risk sets on time of occurrence of the brain cancer. If mobile telephone use changed appreciably over the time scale in which the cases were identified, it might be necessary to retain the matched sets in the analysis, even if the only matching factor were calendar time. Ordinarily, the need to take the matching into account in the analysis is evaluated by comparing the results of an analysis that does take the matching into account with an analysis that ignores the matching. If the results are close, the matching need not be considered further in the analysis.

Another motivation for matching in case-control studies may be to control for variables that would be impossible to control in the analysis without matching. For example, suppose an investigator wishes to control for early-childhood environmental and genetic influences by controlling for family, specifically by using sibling controls. The only practical way to ensure that sibling controls can be used is to select them by matching on sibship during subject ascertainment. Apart from such exceptions, however, the drawbacks of matching typically may outweigh any advantages. If the investigator does decide to match in a case-control study, it may be worth considering using a high matching ratio (the number of controls

matched to each case). A high matching ratio will reduce the probability that any matched set would have completely concordant exposures and reduce the number of matched sets that would be lost to the analysis.

Information Bias

Systematic error in a study can arise because the information collected about or from study subjects is erroneous. Such information is often referred to as *misclassified* if the variable is measured on a categorical scale and the error leads to a person being placed in an incorrect category. For example, a heavy smoker who is categorized as a light smoker is misclassified. Misclassification of subjects can be *differential* or *nondifferential*. Nondifferential misclassification is a misclassification that is unrelated to other study variables. In contrast, with differential misclassification, the misclassification differs according to the value of other study variables. The two key variables to consider with regard to misclassification are exposure and disease.

A common type of information bias is *recall bias*, which occurs in case-control studies in which a subject is interviewed to obtain exposure information after disease has occurred. For example, case-control studies of babies born with birth defects sometimes obtain interview information from mothers after the birth. Mothers who have given birth to a baby with a serious birth defect are thought to be able to recall accurately many exposures during early pregnancy, such as taking nonprescription drugs or experiencing a fever, because the adverse pregnancy outcome serves as a stimulus for the mother to consider potential causes of the birth defect. Mothers of normal babies, however, have had no comparable stimulus to search their memories and may consequently fail to recall exposures such as nonprescription drugs or fevers. The discrepancy in recall gives rise to a particular version of recall bias known as *maternal recall bias*. This problem is distinct from the more general problem of remembering and reporting exposures, which affects all people to some extent and tends to be a nondifferential rather than a differential misclassification.

How can recall bias be prevented? One approach is to frame the questions to aid accurate recall. Improving accuracy of recall reduces recall bias, because it limits the inaccurate recall among controls. Another approach is to use an entirely different control group that will not be subject to the incomplete recall. For example, mothers of babies born with birth defects other than the one under study may provide recall of earlier exposures comparable with that of case mothers. Another approach to avoiding recall bias is to conduct a study that does not use interview information but instead uses information from medical records that was recorded before the birth outcome was known.

Recall bias is a differential misclassification because the exposure information is misclassified differentially for those with or without disease. Although it occurs only in case-control studies, there is an analogous type of differential misclassification that occurs in follow-up studies, in which unexposed people are underdiagnosed for disease more than exposed people. Suppose an investigator conducts a cohort study to assess the effect of tobacco smoking on the occurrence of emphysema. Suppose also that the study asks about medical diagnoses

but does not involve any examinations to check the diagnoses. It may happen that emphysema, a diagnosis that is often missed, is more likely to be diagnosed in smokers than in nonsmokers. Both the smokers and their physicians may be inclined to search more thoroughly for respiratory disease because they are concerned about the effects of smoking. As a result, the diagnosis of emphysema may be missed more frequently among nonsmokers, leading to a differential misclassification of disease. Even if smoking did not lead to emphysema, smokers would appear to have a greater incidence rate of emphysema than nonsmokers because of the greater likelihood that a case of emphysema would remain undiagnosed in a nonsmoker. This bias could be avoided by conducting examinations for emphysema as part of the study itself, thereby avoiding the biased follow-up.

The previous biases are examples of differential misclassification, when the exposure is misclassified differentially according to a person's disease status or disease is misclassified differentially according to a person's exposure status. Differential misclassification can exaggerate or underestimate an effect. A more pervasive type of misclassification, which affects every epidemiologic study to some extent, is nondifferential misclassification. With nondifferential misclassification, exposure or disease (or both) is misclassified, but the misclassification does not depend on a person's status for the other variable. For example, suppose that the study hypothesis concerns the relation between consumption of red wine and the development of emphysema; assume for this example that consumption of red wine is not related to smoking. Unlike the situation for smoking, there is little reason to suppose that those who drink more or less red wine will have a greater or a lesser tendency to be diagnosed with emphysema if they have it. As a result, although some people with emphysema will not have it diagnosed, the proportion of people who do not have their emphysema diagnosed would be expected to be the same for those who do and who do not drink red wine. The underdiagnosis represents some misclassification of emphysema, but because the tendency for underdiagnosis is the same for exposed and unexposed people, the misclassification of disease is nondifferential with respect to exposure. Similarly, if an exposure is misclassified in a way that does not depend on disease status, the exposure misclassification is nondifferential with respect to disease.

Nondifferential misclassification leads to more predictable biases than does differential misclassification. Misclassification of a dichotomous exposure that is nondifferential with respect to disease tends to produce estimates of the effect that are "diluted" or closer to the null or no-effect value than the actual effect. If there is no effect to begin with, nondifferential misclassification of the exposure will not bias the effect estimate.

The simplest case to consider is nondifferential misclassification of an exposure that is measured on a dichotomous scale: exposed versus nonexposed. Suppose that an investigator conducts a case-control study to assess the relation between eating a high-fat diet and subsequent heart attack. Everyone in the study is classified according to some arbitrary cutoff value of dietary fat intake as having a high-fat diet or not. This classification cannot be perfectly accurate because it is almost impossible to avoid some measurement error. In the case of measuring the fat content of a person's diet, there is likely to be substantial error, and some people who do not have a high-fat diet may be classified as having one and vice

versa. If these misclassifications are not related to whether a person gets a heart attack, the misclassification is nondifferential with respect to disease.

The effect of nondifferential misclassification of a dichotomous exposure is illustrated in Table 7-6. On the left are hypothetical data that presume no misclassification with respect to a high-fat diet. The incidence rate ratio (calculated from the odds ratio) is 5.0, indicating a substantially greater mortality rate among those eating a high-fat diet. The center columns show the result if 20% of those who actually do not eat a high-fat diet were inaccurately classified as eating a high-fat diet. This level of misclassification is higher than ordinarily expected, even for an exposure as difficult to measure as diet, but it still involves only a small proportion of the subjects. By moving 20% of those from the No column to the Yes column, the resulting data give a rate ratio of 2.4, less than one half as great as the value with the correct data. In terms of the effect part of the risk ratio, the excess risk ratio of 4.0 ($= 5.0 - 1$) has been reduced to 1.4 ($= 2.4 - 1$), which means that about two thirds of the effect has been obscured. Notice that we have transferred both 20% of cases and 20% of controls. Nondifferential misclassification of exposure implies that these percentages of misclassified subjects among cases and controls will be equal. If the proportions of cases and controls that were misclassified differed from one another, the misclassification would be differential with respect to disease.

The third set of columns in Table 7-6 adds further nondifferential misclassification: 20% of cases and controls who ate a high-fat diet are misclassified as not having a high-fat diet. This misclassification is added to the misclassification in the other direction, with the result as shown in the right part of the table. With this additional misclassification, the rate ratio has declined to 2.0, even closer to the null value of 1.0, nullifying three fourths of the effect seen in the correctly classified data.

Nondifferential misclassification of a dichotomous exposure will always bias an effect, if there is one, toward the null value. If the exposure is not dichotomous, there may be bias toward the null value, but there may also be bias away from the null value, depending on the categories to which individuals get misclassified.

Table 7-6 NONDIFFERENTIAL MISCLASSIFICATION IN A HYPOTHETICAL CASE-CONTROL STUDY

	Correct Classification		Nondifferential Misclassification			
			20% of No \rightarrow Yes		20% of Yes \rightarrow No	
	High-Fat Diet		High-Fat Diet		High-Fat Diet	
	No	Yes	No	Yes	No	Yes
Myocardial infarction cases	450	250	360	340	410	290
Controls	900	100	720	280	740	260
	RR = 5.0		RR = 2.4		RR = 2.0	

Nondifferential misclassification between two exposure categories usually makes the effect estimates for those two categories converge toward one another.⁴

Confounding

Confounding is a central issue for epidemiologic study design. A simple definition of confounding is the *confusion of effects*. This definition implies that the effect of the exposure is mixed with the effect of another variable, leading to a bias. Consider a classic example: the relation between birth order and the occurrence of Down syndrome. Figure 7-3 shows data on birth order and Down syndrome from the work of Stark and Mantel.⁵

These data show a striking trend in prevalence of Down syndrome with increasing birth order, which can be described as the effect of birth order on the occurrence of Down syndrome. The effect of birth order, however, is a blend of whatever effect birth order has by itself and the effect of another variable that is closely correlated with birth order: the age of the mother. Figure 7-4 gives the relation between mother's age and occurrence of Down syndrome from the same data. It indicates a much stronger relation between mother's age and Down syndrome. In Figure 7-3, the prevalence increased from about 0.6/1000 at the first birth to 1.7/1000 for birth order of 5 or greater, a respectably strong trend. In Figure 7-4, however, the prevalence increases from 0.2/1000 at the youngest category of mother's age to 8.5/1000 at the highest category of mother's age, more than a 40-fold increase. (The vertical scale changes from Figure 7-3 to Figure 7-4.)

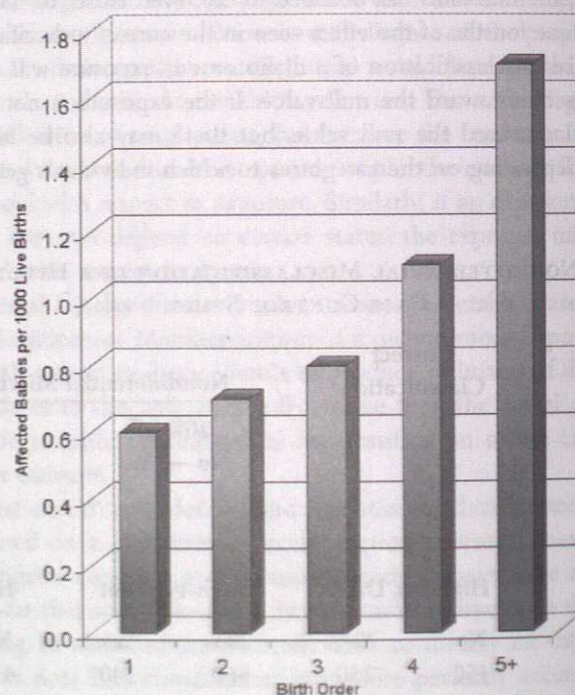


Figure 7-3 Prevalence of Down syndrome at birth by birth order. (Data from Stark and Mantel.⁵)

CONFOUNDING BY INDICATION

Pharmacoepidemiologists study the epidemiology of intended and unintended drug effects, often by using nonexperimental studies. In these studies, the essential comparisons involve a contrast of outcomes for individuals who have taken a specific drug with those who have not taken the drug. Without a randomized trial, it can be challenging to design a study that yields a valid comparison of drug takers with nontakers. The main challenge comes from a phenomenon that epidemiologists refer to as *confounding by indication*. The problem arises from the fact that those who take a drug usually differ from those who do not according to the medical indication for which the drug was prescribed. Even if the comparison group represents patients with the same disease who received a different therapy or none at all, there typically are differences in disease severity or other risk factors between populations who receive different treatments. These differences introduce a bias in the comparison that is called confounding by indication, which is described further in Chapter 13.

Because birth order and the age of the mother are highly correlated, we can expect that the mothers who are giving birth to their fifth baby are, as a group, considerably older than mothers who are giving birth to their first baby. Therefore the comparison of high-birth-order babies with lower-birth-order babies is to some extent a comparison of babies born to older mothers with babies born to younger mothers. Thus, the birth-order comparison in Figure 7-3 mixes the effect of mother's age with the effect of birth order. The extent of the mixing depends on the extent to which mother's age is related to birth order. This mixing of effects is called *confounding*; the birth order effect depicted in Figure 7-3 is confounded by the effect of mother's age.

Is the effect of mother's age in Figure 7-4 also confounded by the effect of birth order? This is a reasonable question; the answer depends on whether birth order has any effect at all on its own. Because the effect in Figure 7-4 for mother's age is so much stronger than the effect in Figure 7-3 for birth order, we know that birth order cannot fully explain the maternal age effect, whereas it remains a possibility that maternal age fully accounts for the apparent effect of birth order. A good way to resolve the extent to which one variable's effect explains the apparent effect of the other is to examine both effects simultaneously. Figure 7-5 presents the prevalences of Down syndrome at birth by both birth order and mother's age simultaneously.

Figure 7-5 shows that within each category of birth order, looking from the front to the back, there is the same striking trend in prevalence of Down syndrome with increasing maternal age. In contrast, within each category of maternal age, looking from left to right, there is no discernible trend with birth order. Thus, the apparent trend with birth order in Figure 7-3 is entirely explained by confounding by maternal age. There is no confounding in the other direction: Birth order does not confound the maternal age association, because birth order has no effect. We call the apparent effect of birth order in Figure 7-3 the *crude*

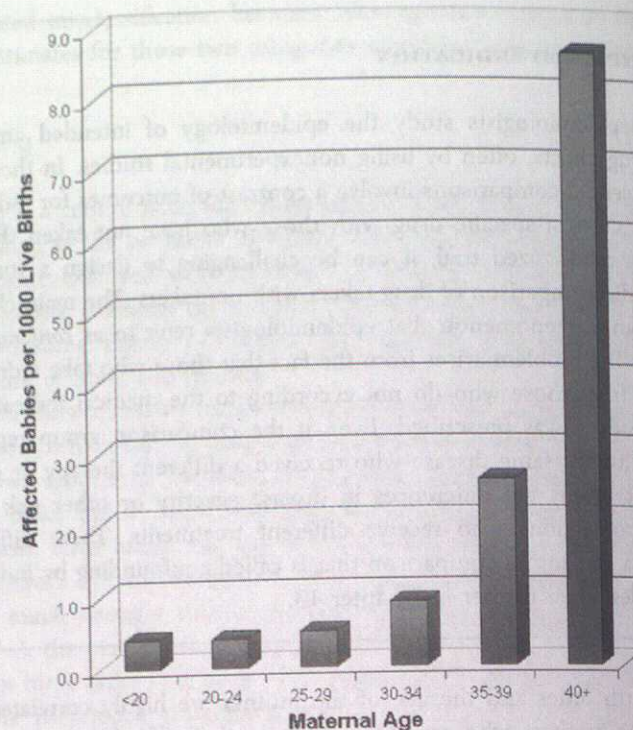


Figure 7-4 Prevalence of Down syndrome at birth by mother's age.

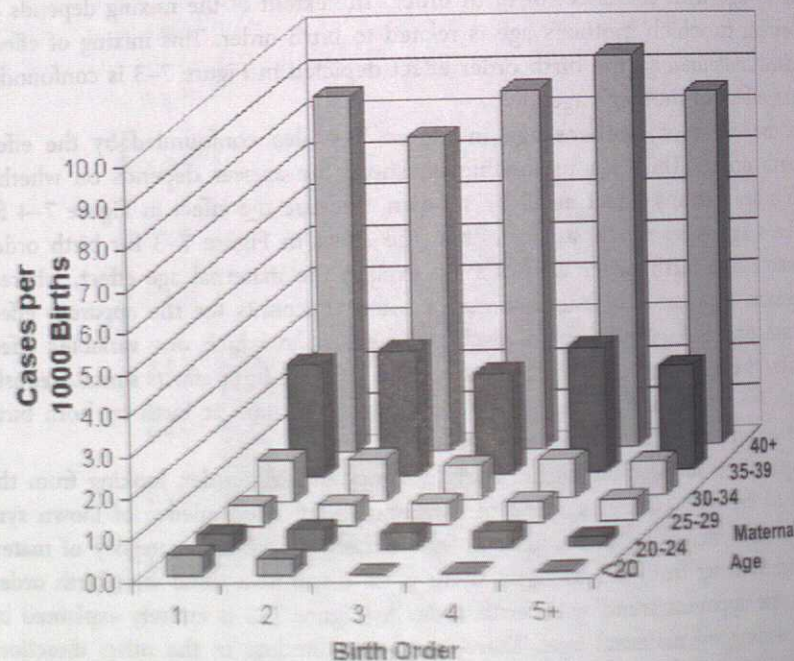


Figure 7-5 Prevalence of Down syndrome at birth by birth order and mother's age.

effect of birth order. In Figure 7-5, we can see that within categories of maternal age there is no birth order effect, and the crude effect in this instance is entirely a result of confounding.

Although the maternal age effect in Figure 7-4 is not confounded by birth order, which appears to have no effect on its own, it is confounded by other factors. We can be sure of that because age is just a marker of time. There must be biologic events that occur during a woman's aging process that lead to the sharp increase in occurrence of Down syndrome among the offspring of older mothers. Mother's age is thus a proxy for unidentified events that more directly account for the occurrence of Down syndrome. When these events are identified, we may ultimately find that mother's age has no effect after we take into account the biologic changes that are correlated with age. In this sense, we can say that the apparent effect of mother's age is presumably confounded by unknown factors.

The research process of learning about and controlling for confounding can be thought of as a walk through a maze toward a central goal. The path through the maze eventually permits the scientist to penetrate deeper levels of understanding. In this example, the apparent relation between Down syndrome and birth order can be explained entirely by the effect of mother's age, but that effect ultimately will be explained by other factors that have not yet been identified. As the layers of confounding are left behind, we gradually approach a deeper causal understanding of the underlying biology. Unlike a maze, however, this journey toward the goal of biologic understanding does not have a clear end point, because there is always room to understand the biology in a deeper way.

Confounding previously was defined as the confusion of or mixing of effects. Strictly speaking, the exposure variable may or may not have an effect; in the Down syndrome example, birth order did not have an effect. The confounding variable, however, must have an effect on the outcome to be confounding. Theoretically, a confounding variable should be a cause of the disease, but in practice, it may be only a proxy or a marker for a cause. That is the case for mother's age, which by itself does not cause Down syndrome but serves as a marker for unknown biologic events that accumulate with time. Whether an actual cause or just a marker for a cause, a confounder is a predictor of disease occurrence.

Nevertheless, not every predictor of disease occurrence is a confounding factor. For confounding to occur, a predictor of disease occurrence must also be imbalanced across exposure categories. Suppose that age is a risk factor for a given disease (as it usually is). Age would not be confounding unless the age distributions of people in the various exposure categories differed, as they did for smoking and nonsmoking women in Table 1-2. If every exposure category contains people whose age distribution is the same as that for people in other exposure categories, the comparison of disease rates across exposure categories is not distorted by age differences. On the other hand, if age is imbalanced across exposure categories, the comparison of one exposure category with another involves the comparison of people whose age distributions differ. Under those circumstances, the effect of exposure will be confounded with the effect of age to an extent that depends on the strength of the relation between age and the disease and on the extent of the age imbalance across exposure categories.

Table 7-7 DEATHS AMONG PATIENTS WHO RECEIVED TOLBUTAMIDE AND PLACEBO IN THE UNIVERSITY GROUP DIABETES PROGRAM IN 1970

	Tolbutamide	Placebo
Deaths	30	21
Surviving	174	184
Total	204	205
Mortality proportion	0.147	0.102

Consider another example. In 1970, the University Group Diabetes Program published the results of a randomized trial designed to assess how well three treatments for diabetes prevented fatal complications.⁶ Table 7-7 presents the crude data comparing one of the treatments, the drug tolbutamide, with placebo, with respect to total mortality over a period that averaged 7 years.

The proportion of subjects who died was greater in the tolbutamide group than in the placebo group, a surprising result that spurred a long and bitter controversy and brought tremendous scrutiny to these study results. If we measure the effect of treatment as the difference in proportion of those who died in the tolbutamide and placebo groups, we estimate an adverse effect of tolbutamide of $0.147 - 0.102 = 0.045$. This result translates to an estimate that subjects who receive tolbutamide face an additional risk of 4.5% of dying over 7 years compared with subjects receiving placebo.

Although this study was a randomized experiment, the random assignment in this case led to imbalances between the tolbutamide and placebo groups with respect to age. Randomization is intended to balance potential confounding factors between the compared groups, but it cannot guarantee such a balance. In this case, the tolbutamide group comprised subjects who were older on average than the placebo group. Because age is strongly related to the risk of death, this imbalance in age introduced confounding. In the Down syndrome example, we removed the confounding by examining the effect of birth order within categories of mother's age. This process is called *stratification*. We can also stratify the data from the University Group Diabetes Program by age (Table 7-8).

Table 7-8 shows that of the 204 subjects who received tolbutamide in the study, 98 (48%) were 55 years old or older; in contrast, only 85 of 205 placebo subjects (41%) were age 55 or older. This difference may not appear striking,

Table 7-8 DEATHS AMONG SUBJECTS WHO RECEIVED TOLBUTAMIDE AND PLACEBO IN THE UNIVERSITY GROUP DIABETES PROGRAM IN 1970, STRATIFIED BY AGE

	Age < 55		Age 55+	
	Tolbutamide	Placebo	Tolbutamide	Placebo
Dead	8	5	22	16
Surviving	98	115	76	69
Total	106	120	98	85
Mortality proportion	0.076	0.042	0.224	0.188
Difference in proportion	0.034		0.036	

but the difference in the risk of dying during the 7 years is strikingly greater for those age 55 or older than for younger subjects. With age so strongly related to the risk of death, the difference in the age distribution is potentially worrisome. Did it lead to confounding by age? To answer that, we can look at the difference in the proportion who died, comparing tolbutamide with placebo, in each of the two age groups. In both groups, there is an approximately 3.5% greater risk of death over the 7 years for the tolbutamide group than for the placebo group. (The data show a difference of 3.4% for the younger group and 3.6% for the older group. As a summary measure, we can average these two values and call the overall difference 3.5%. Technically, we would want to take an average that weighted each of the age categories according to the amount of data in that category.) When age was ignored, we found a difference of 4.5%. The value of 4.5% that we obtained from the crude data is confounded and gives an overestimate of the adverse effect of tolbutamide. The value of 3.5% obtained after the age stratification may not be completely unconfounded by age. Because we used only two age categories, it is possible that age differences remain within the age categories in Table 7-8. Nevertheless, even with this simple age stratification, the estimate of effect is lower than the estimate from the crude data. The crude data overestimate the adverse effect of tolbutamide by almost 30% (4.5% is almost 30% greater than 3.5%). The topic of stratification is discussed further in Chapter 10.

PROPERTIES OF A CONFOUNDING FACTOR

Confounding can be thought of as a mixing of effects. A confounding factor therefore must have an effect, and it must be imbalanced between the exposure groups that are being compared. These conditions imply that a confounding factor must have two associations:

- A confounder must be associated with the disease (either as a cause or as a proxy for a cause, but not as an effect of the disease).
- A confounder must be associated with the exposure.

There is also a third requirement. A factor that is an effect of the exposure and is an intermediate step in the causal pathway from exposure to disease will have the previously described properties, but causal intermediates are not confounders; they are part of the effect that we wish to study. For example, if a diet high in saturated fat leads to higher levels of low-density lipoproteins (LDL) in the blood, and a high LDL level leads to atherosclerosis, a high LDL level will be associated with both diet and atherosclerosis. Nevertheless a high LDL level does not confound the relation between diet and atherosclerosis; it is part of the exposure's effect and should not be considered confounding. Any effect of the exposure, whether it is part of the causal pathway to the disease or not, is not a confounder. Thus, the third property of a confounder is the following:

- A confounder must not be an effect of the exposure.

The University Group Diabetes Program illustrates that even randomization cannot prevent confounding in all instances. In this case, it led to an age imbalance, which caused a moderate amount of confounding. The confounding in the Down syndrome example was greater, in large part because the association between mother's age and birth order is stronger than the association between age and tolbutamide that the randomization produced in the University Group Diabetes Program.

Notice that confounding can cause a bias in either direction. It can cause an overestimate of the effect, as the confounder mother's age did for birth order and Down syndrome and the confounder age did for tolbutamide and death, or it can cause an underestimate of an effect, as the confounder age did for smoking and death in the example in Chapter 1. The bias introduced by confounding occasionally can be strong enough to reverse the apparent direction of an effect,⁷ as illustrated in Chapter 1 when comparing the death rates in Panama and Sweden.

CONTROL OF CONFOUNDING

Confounding is a systematic error that investigators aim to prevent or to remove from a study. There are three methods that are used to prevent confounding. One of them, *randomization*, or the random assignment of subjects to experimental groups, can be used only in experiments. The second method, *restriction*, involves selecting subjects for a study who all have the same value or almost the same value for a variable that would otherwise be a confounding variable. Restriction can be used in any epidemiologic study, regardless of whether it is an experiment or not. The third approach is *matching*, which is an effective way to prevent confounding in cohort studies, but as discussed earlier, causes a selection bias in case-control studies. Because no method prevents confounding completely, these design methods may be best viewed as methods to limit confounding.

In experiments, in which the investigator assigns the exposure to study subjects, randomization confers powerful benefits. With a sufficiently large study population, randomization produces two or more study groups with almost the same distribution of characteristics. This similarity for all variables implies that the compared groups will be similar for risk factors that predict the outcome of interest and that these risk predictors therefore will not confound. Randomization cannot guarantee the absence of confounding; a random process can still lead to confounding imbalances, such as the age imbalance that occurred in the University Group Diabetes Program experiment and shown in Tables 7-7 and 7-8. The likelihood of a large imbalance, however, becomes small as the number of subjects who are randomized increases. Perhaps the most important benefit of randomization is that it prevents confounding for unidentified factors as well as for factors that are already known to be of concern. Even unknown risk factors will not confound a randomized experiment of sufficient size.

Restriction, unlike randomization, cannot control for unknown confounding factors, but it is more certain to prevent confounding for those factors for which it is employed. For example, in a study of alcohol drinking and cancer of the throat, smoking may be considered a likely confounding variable. Smoking is a cause of throat cancer, and people who drink alcohol smoke more than people who do not drink alcohol. If the study were confined to nonsmokers, smoking

IS CONFOUNDING IN A RANDOMIZED EXPERIMENT A BIAS?

Earlier in this chapter, I proposed that if an error in a study would decrease if the study were larger, then that error is a random error, whereas an error that would not decrease if the study were larger is a systematic error. Confounding is usually considered a systematic error, but confounding in an experiment is an exception. In all types of epidemiologic studies, confounding arises from imbalances in risk factors for the outcome across the exposure categories. Uniquely in randomized experiments, however, these imbalances are determined by random assignment. As a result of the law of large numbers, the larger the experiment, the more closely the randomly assigned groups will resemble one another in their distributions of risk factors. Because the amount of confounding depends on the size of the experiment, confounding in an experiment is an example of random error rather than systematic error. For systematic errors, replicating the study replicates the error, but for confounding in an experiment, replicating the study (with a new random assignment) will not replicate the same confounding because there will be an entirely new set of assignments to the study groups. Despite being an example of random error rather than systematic error, confounding in an experiment can be controlled using the same methods to control confounding in nonexperimental studies.

In this discussion, a large experiment does not necessarily mean one with many participants; rather, it is one that has a large number of random assignments. For example, a study may involve the random assignment of a community intervention to eight cities that contain millions of people. With only eight random assignments, however, it is not large enough to prevent substantial confounding.

could not be confounding. Similarly, if age is thought to be a likely confounding factor in a study, confounding by age can be prevented by enrolling subjects who are all the same age. If everyone in a study has the same value for a variable, that factor can no longer vary in the study setting; it becomes a constant. For confounding to occur, a confounding factor must be associated with exposure, but if a factor is constant, it cannot be associated with anything. Restriction is an effective way to prevent confounding in any study.

Restriction is used in experiments in addition to randomization to be certain that confounding for certain factors does not occur. It is also used by laboratory scientists conducting animal experiments to prevent confounding and enhance the validity of their studies. Typically, a researcher conducting an experiment with mice seeks only mice bred from the same laboratory and that have the same genotype, the same age, and sometimes the same sex.

It may appear puzzling that restriction is not used more often in epidemiologic research. One explanation is that many researchers have been taught that an epidemiologic study, whether an experiment or a nonexperimental study, should comprise study subjects whose characteristics are representative of the target population for whom the study results are intended. The goal of *representativeness* appears to

work contrary to the method of restriction, which provides a study population that is homogeneous and therefore not similar to most target populations of interest. Elevating the importance of representativeness is a fallacy that has plagued epidemiologic studies for decades. As explained in Chapter 3, the notion that representativeness is a worthwhile goal presumably stems from the arena of survey research, in which a sample of a larger population is surveyed to avoid the expense and trouble of surveying the entire population. The statistical inference that such sampling allows is only superficially similar to the scientific inference that is the goal of epidemiologic research. For scientific inference, the goal is not to infer a conclusion that would apply to a specific target population but rather to infer an abstract theory that is not tied to a specific population. It is possible to make such a scientific inference more readily without confounding; restriction enhances the ability to make a scientific inference, as those who work with laboratory animals know.

What about the concern that restriction makes it difficult to know whether a studied relation applies to people with characteristics different from those in a study population? For example, suppose that an investigator uses restriction to study the effect of drinking wine on cardiovascular disease risk among people who are 60 years old. Would the study results apply to people who are 45-years old? The answer is maybe; without outside knowledge, it is not possible to say whether the study results apply to people who are 45 years old. This uncertainty leaves open the possibility of an erroneous or incomplete conclusion, but such is the nature of science. It is nevertheless wrong to think that the theorization needed to apply the results of a study to people with different characteristics could be replaced by mechanical sampling. If an investigator suspects that the effect of wine consumption on cardiovascular risk is different for 60-year-olds and 45-year-olds, he or she would want to select a group of 45-year-olds to study in addition to 60-year-olds. The number of the study subjects and their age distribution should not reflect the age distribution in some target population; why let the demographics of a locale dictate the age distribution of subjects that are chosen for study? Instead, the investigator can choose to study subjects of whatever age seems interesting and in numbers that suit the study design rather than reflect the numbers of people in a target population at those ages. Scientifically, there is no specific target population. There is instead a scientific theory about wine, cardiovascular disease risk, and perhaps age. The theory is the real target of inference. A valid study is the best route to a correct inference, and restriction, rather than representativeness, is the more desirable means to achieve the correct inference.

Matching should be distinguished from restriction. With restriction, all subjects are confined to a single value or narrow range of values for one or more factors that are suspected of being possible confounding factors. Matching imposes no constraint on the index subjects, those who are the target of the matching. The other subjects are selected to conform to the index series for whatever matching factors are employed. Suppose one is conducting a cohort study and wishes to control for age by matching. Generally, when matching in a cohort study, the index series is the exposed series, and the goal of matching is to assemble an unexposed series that has the same age distribution as the exposed subjects. There are two ways to accomplish this goal. One approach is to describe the age distribution of exposed subjects and then select unexposed subjects to replicate that age distribution. That approach is called *frequency matching*. The other approach is to

take the exposed subjects one by one and to find for each of them an unexposed subject that has a matching age. The investigator also can select two or three or any fixed number of unexposed subjects to match with each exposed subject. This approach is called *individual matching*. Whether frequency matching or individual matching is used, the result is that for the matched factor, the exposed and unexposed cohorts will have the same distribution, and therefore that factor will not be confounding. To the extent that the matching is loose rather than tight, such as matching within 5 years of age rather than match to the exact year of age, there may still be minor differences in the age distribution. The tighter the match, the more effective the elimination of age confounding. For variables that are categorical, such as sex, defining a tight match is not an issue.

Matching is very effective in preventing confounding, but there are a few cautions to consider. The greatest caution, described earlier in this chapter, is that it does not work to prevent bias in case-control studies. In cohort studies, it does work well, and it can lead to results that are unbiased by the matching factors, provided that the cohort is followed for a short enough time for the matching to be maintained. Because matching is accomplished for people, not person-time, as people are lost to follow-up for various reasons, the initially equal distributions across cohorts for matching factors may become different if exposed people and unexposed people are lost to the study at different rates.

Matching can be an expensive process. To match an unexposed cohort to a sizable exposed cohort can be costly and consequently has seldom been attempted. The main exception is when all potential subjects and their data are already stored in a data warehouse or database. In that case, matching an unexposed cohort is no more costly than implementing a computer program to find the matching subjects, and matching becomes considerably more cost-effective. There is a drawback to using matching within a database, however. If the data are already available in a database, matching will result in excluding some subjects from the study. Some excluded subjects will be possible matches that were not used because there were closer matches that were chosen instead; others may be outliers that are unmatchable. This loss of subjects may lessen the appeal of matching compared with alternative methods that can be employed in the data analysis that retain all possible subjects. On the other hand, the exclusion of unmatchable subjects may enhance the validity of the study and be preferable to including them; this point is discussed further in Chapter 12.

Control of confounding in the data analysis requires that the study data include adequate information about the confounding factor or factors. Two methods can be used to deal with confounding in the data analysis. One is stratification, a technique that was illustrated in Table 7-8 and in Chapter 1. It is discussed in greater detail in Chapter 10. The other approach is by using regression models, an analytic technique that is described in Chapter 12.

QUESTIONS

1. Suppose a case-control study could be expanded to be infinitely large. Which sources of error would be eliminated by such a study, and which would not? Suppose that a randomized trial could be infinitely large. Which sources of error would remain in such a trial?

2. Will a larger study have less bias than a smaller study? Why or why not?
3. When recall bias occurs, patients who have been afflicted with a medical problem, such as a heart attack, give responses about possible causes of that problem that differ from those given by nonafflicted subjects. Whose responses are thought to be more accurate?
4. Suppose that in analyzing the data from an epidemiologic study, a computer coding error led to the exposed group being classified as unexposed and the unexposed group being classified as exposed. What specific effect would this error have on the reported results? Is this a bias? If so, what type? If not, what type of error is it?
5. Explain the difference between a confounding factor and a potential confounding factor. In what situations might a potential confounding factor not end up being a confounding factor?
6. The incidence rate of cardiovascular disease increases with increasing age. Does that mean that age always confounds studies of cardiovascular disease in the same direction? Why or why not?
7. The effectiveness of randomization in controlling confounding depends on the size of the experiment. Consider an experiment to study the effect of nutritional education of schoolchildren on their serum cholesterol levels. Suppose that the study involved randomly assigning 10 classrooms with 30 children each to receive a new curriculum and assigning another 10 classrooms with 30 children each to receive the old curriculum. Should this be considered a study that compares two groups with 300 in each group or 10 in each group from the viewpoint of the effectiveness of controlling confounding by randomization?
8. Confounding by indication arises because those who take a given drug differ for medical reasons from those who do not take the drug. Is this problem truly confounding, or is it more appropriately described as a selection bias?
9. Those who favor representative studies claim that one should not generalize a study to a population whose characteristics differ from those of the study population. A study of smoking and lung cancer in men would tell nothing about the relation between smoking and lung cancer in women. Give the counterarguments. (Hint: if the study were conducted in London, would the results apply to those who lived in Paris?)

REFERENCES

1. Nichol KL, Nordin JD, Nelson DB, Mullooly JP, Hak E. Effectiveness of influenza vaccine in the community-dwelling elderly. *N Engl J Med*. 2007;357:1373-1381.

2. Simonsen L, Reichert TA, Viboud C, Blackwelder WC, Taylor RJ, Miller MA. Impact of influenza vaccination on seasonal mortality in the US elderly population. *Arch Intern Med*. 2005;165:265-272.
3. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *Int J Epidemiol*. 2006;35:337-344.
4. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
5. Stark CR, Mantel N. Effects of maternal age and birth order on the risk of mongolism and leukemia. *J Nat Cancer Inst*. 1966;37:687-698.
6. University Group Diabetes Program. A study of the effects of hypoglycemic agents on vascular complications in patients with adult onset diabetes. *Diabetes*. 1970;19(suppl 2):747-830.
7. Reintjes R, de Boer A, van Pelt W, Mintjes-de Groot J. Simpson's paradox: an example from hospital epidemiology. *Epidemiology*. 2000;11:81-83.