11

Analysis of Epidemiologic Studies: Evaluating the Role of Bias

Bias may be defined as any systematic error in an epidemiologic study that results in an incorrect estimate of the association between exposure and risk of disease. Since epidemiologic studies involve free-living human beings, even the most rigorously designed investigation will have the potential for one or more types of bias arising from a large number of specific sources including the manner in which subjects are selected into the study and the way in which information is obtained, reported, or interpreted. Consequently, evaluating the role of bias as an alternative explanation for an observed association is a necessary step in interpreting any study result. Unlike chance and confounding, which can be evaluated quantitatively, the effects of bias are far more difficult to evaluate and may even be impossible to take into account in the analysis. For this reason, it is of paramount importance to design and conduct each study in such a way that every possibility for introducing bias has been anticipated and that steps have been taken to minimize its occurrence. Since such errors may, nonetheless, occur, it is also important in the interpretation of findings to consider, at the very least, the types of bias that could have arisen in that particular study as well as the most likely direction and magnitude of their impact.

In this chapter, we provide a brief overview of various types of bias, some practical means to minimize their occurrence, as well as ways to evaluate their influence on the estimates of effect from a study. The reader should also refer to Chapters 6 to 8 for detailed discussions of the particular types of bias of most concern for each analytic study design.

TYPES OF BIAS

There are a number of ways of categorizing and naming the different types of bias that may distort the association between exposure and disease observed in a particular study [11]. One simple but useful approach

involves two general classes of systematic error under which the specific types can fall. The first, selection bias, refers to any error that arises in the process of identifying the study populations. The second general category, observation or information bias, includes any systematic error in the measurement of information on exposure or outcome.

Selection Bias

Selection bias can occur whenever the identification of individual subjects for inclusion into the study on the basis of either exposure (cohort) or disease (case-control) status depends in some way on the other axis of interest. In other words, if in a case-control study, selection of cases and controls is based on different criteria and these, in turn, are related to exposure status, bias will result. Similarly, in a cohort study, if the choice of particular exposed and nonexposed individuals is related to their development of the outcome of interest, selection bias can occur. Selection bias is a particular problem in case-control and retrospective cohort studies, where both the exposure and outcome have occurred at the time individuals are selected into the study. Selection bias is unlikely to occur in a prospective cohort study because exposure is ascertained before the development of any outcomes of interest. In all instances where selection bias does occur the result is an observed relation between exposure and disease that is different among those who are entered into the study than among those who would have been eligible but were not chosen to participate.

Selection bias can result from a number of circumstances related to the way in which individuals are ascertained and selected for study. These include factors such as differential surveillance, diagnosis, or referral of individuals into the study. One classic example of this type of bias relates to case-control investigations of oral contraceptives (OC) and thromboembolism. When the first hospital-based case-control studies of OC use and thromboembolism were reported, there was some concern that since physicians were already aware of a possible relationship of these agents with thromboembolism, a proportion of the women in the case series had been hospitalized for evaluation of this disease because they were currently using OCs [12]. If so, any increased frequency of current OC use among women hospitalized for thromboembolism might actually be due, at least in part, to the fact that hospitalization and the determination of the diagnosis were both influenced by a history of OC use. Similarly, criticism concerning the increased risk of uterine cancer associated with use of exogenous estrogens observed in early case-control studies of this hypothesis was based on the argument that women who used estrogens experienced uterine bleeding and were, therefore, more likely to seek medical attention and have a diagnostic evaluation than women not on estrogens [5]. In such circumstances, further evi-

dence supporting the validity of each of the observed associations could be derived from examination of the association among women who met certain objective diagnostic criteria and whose disease would consequently have come to medical attention regardless of exposure status [6].

Another source of selection bias in case-control studies derives from refusal or nonresponse among either study group. While neither low participation rates nor different rates of response between cases and controls necessarily indicates the presence of bias, if the rates of response are also related to exposure status, then bias will be a reasonable alternative explanation for any observed association between exposure and disease. For example, if controls are selected through household survey, it is very possible that nonresponse will be related to a large number of demographic and lifestyle variables associated with employment, some or all of which may also be risk factors for the outcome of interest. Under such circumstances, the possibility of selection bias will be a major problem in interpreting study results.

Observation (Information) Bias

Observation or information bias results from systematic differences in the way data on exposure or outcome are obtained from the various study groups. If data are inaccurate or incomplete, spurious associations may be introduced only if the inaccuracy or incompleteness affects the two groups to an unequal degree. There are several specific types of observation bias, depending on the source of noncomparability.

Recall bias arises when individuals with a particular adverse health outcome remember and report their previous exposure experience differently from those who are not similarly affected, or when those who have been exposed to a potential hazard report subsequent events with a different degree of completeness or accuracy than those nonexposed. This type of bias is especially problematic in case-control and retrospective cohort studies, since both exposure and disease have already occurred at the time participants enter into the study. One of the most common methods of gathering information, particularly in a case-control investigation, is by interviewing either the study subjects themselves or some surrogate, such as spouses of participants or mothers of affected children. Individuals who have experienced a disease or adverse health outcome tend to think about the possible "causes" of their illness and are likely to remember their exposure histories differently from those unaffected by the disease. This bias may also result from the selective recall of families of affected individuals as compared with healthy subjects. Recall bias can lead to either an over- or underestimate of the association between exposure and disease, depending on whether the cases recall their exposure to a greater or lesser extent than the controls.

A second type of systematic error in collecting information is interviewer bias. This refers to any systematic difference in the soliciting, recording, or interpreting of information from study participants and can affect every type of epidemiologic study. In case-control studies, this is a particular concern with respect to the ascertainment of exposure, since knowledge of a subject's disease status may result in differential probing for previous exposure history. Bias in ascertainment of exposure may, for the same reason, also be a problem in retrospective cohort studies. For prospective cohort studies, however, this will not be an issue because the outcome of interest has not yet occurred at the time exposure status is determined. On the other hand, there is particular potential for observation bias in the assessment of outcome for both retrospective and prospective cohort studies. Since information concerning a subject's exposure status is available at the time disease status is determined, investigators who are aware of the study hypothesis may be more (or less) likely to record the outcome of interest for individuals known to have the exposure under examination. This type of bias can also affect the ascertainment of outcome in intervention studies that do not utilize placebo control to maintain observer blindness. In all these circumstances, an association between the exposure and disease might result from or be obscured by this type of systematic error.

The major source of bias in cohort studies concerns the potential for loss of subjects to follow-up due to the necessity of following individuals for a period of time after exposure to determine whether they develop the outcome of interest. When persons lost to follow-up differ from those who remain with respect to both the exposure and the outcome, any observed association will be biased. For example, in a cohort study using mailed questionnaires to evaluate the relation between smoking and myocardial infarction (MI), to the extent that those who both smoke and develop MI are less (or more) likely to respond than nonsmokers who develop the disease, a biased estimate of the exposure and outcome will be obtained. The potential for bias due to losses to follow-up is present no matter how small the proportion of loss is, as long as such loss is related to both exposure and disease.

Another major type of observation or information bias is misclassification, which occurs whenever subjects are erroneously categorized with respect to either exposure or disease status. Since in any study some degree of inaccuracy in reporting or recording information is inevitable, misclassification is always a potential concern. The effect of such misclassification depends on whether the misclassification with respect to exposure (or disease) is dependent on the individual's disease (or exposure) status. When the misclassification is random or nondifferential, the proportions of subjects erroneously classified in the study groups are approximately equal. For example, in a case-control study of spermicides and birth defects [7], the investigators reported increased risks of

several types of congenital malformations among women identified as having filled a prescription for these preparations within 600 days prior to the birth of the child. While defining exposure in this way will inevitably result in some degree of misclassification of the true exposure, the actual use of spermicides during a period relevant to teratogenicity, such misclassification is likely to be random, since recording of prescription data was made before the birth of the child. Because random misclassification increases the similarity between the exposed and nonexposed groups, any true association between the exposure and disease will be diluted.

Some degree of random misclassification of exposure and disease is present in almost all types of epidemiologic studies. Retrospective cohort studies of occupational exposures, for example, often obtain exposure information from records compiled many years prior to the initiation of the study. In addition, they frequently must use variables such as work assignment or membership in a union or professional society as a surrogate for exposure to a particular hazard. These proxy variables are, at best, only crude markers of actual exposure level, and at worst may bear little resemblance to any individual subject's true status. It is, however, unlikely that the accuracy or completeness of these records would be different for those who developed the outcome of interest and those who did not. Similarly, studies utilizing only self-reported exposure may be subject to substantial amounts of misclassification, depending on the nature of the study population and the particular exposures. Again, as long as the misclassification is random, it can only serve to dilute any true association between the exposure and outcome.

A more serious problem arises if the proportions of subjects misclassified differ between the study groups. The effect of such nonrandom or differential misclassification is that the observed estimate of effect can be biased in the direction of producing either an overestimate or underestimate of the true association, depending on the particular situation. It is even possible that the estimate will be correct due to the play of chance or the effect of an additional bias. Unfortunately, it is often difficult if not impossible to estimate the precise effect of differential misclassification.

CONTROL OF BIAS: STUDY DESIGN

The prevention and control of potential biases must be accomplished largely through careful study design. For some types of bias, an evaluation and control of their effects can be accomplished, at least to a partial degree, in the analysis. Others, however, such as selection bias, may not be rectifiable once they have occurred. Consequently, prevention of bias

in the design phase of an investigation is crucial to the validity of the study results. There are a number of design features that can minimize the potential for bias in any study, ranging from general considerations, such as the choice of a study population and the sources of data to be utilized, to specific features of the data collection process.

Choice of Study Population

There are many ways in which the choice of study population can minimize potential biases. For example, the selection of hospitalized controls in a case-control study will increase comparability with the cases in terms of willingness to participate, the presence of selective factors that influenced the subjects' choice of a particular hospital, and awareness of antecedent exposures and events. This will decrease the likelihood of nonresponse, selection, and recall bias. For cohort studies and clinical trials, in which the ability to locate individuals over a period of time is crucial to minimizing bias due to losses to follow-up, investigators frequently choose populations that are well defined with respect to occupation, place of employment, area of residence, or some other characteristic that gains access to centralized sources of information, such as alumni of a particular institution, U.S. veterans, or members of a health maintenance organization. Another factor often considered in choosing a study population that will minimize biases related to nonresponse or losses to follow-up, particularly in intervention studies, is the selection of a population at above average risk of developing the outcome under investigation. Such individuals are likely to be more interested in taking part in health research than those at usual risk of the disease and thus are likely to maintain a higher level of compliance and commitment to the study.

Methods of Data Collection

In any analytic study, the actual means by which data are collected will have a major impact on the validity of results. There are usually many ways to obtain the same piece of information. The particular method implemented in any given investigation may mean the difference between obtaining useful and informative data or uninterpretable results. From a practical standpoint, two major aspects of the design of data collection procedures can minimize bias: (1) the construction of the specific instrument(s) to obtain information, including questionnaires, interviews, physical examinations, and forms for abstracting data from records; and (2) the administering of those instruments by study personnel. It is crucial to keep very clearly in mind that whatever the sources

and methods of data collection utilized in a study, they should be similar for all study groups.

With respect to the data collection instruments, one of the most important ways to minimize bias is the use of highly objective, closed-ended questions. For example, if the variable of interest in a particular study were blood pressure level, information could be obtained through a variety of means ranging from a question about history of hypertension on a self-administered questionnaire to calculating the average of several blood pressure readings taken by a trained observer using a standardized protocol. A question about history of hypertension is clearly the most subjective and, therefore, has the greatest potential for introducing bias. Obtaining information on last recorded blood pressure, while not requiring a judgment about what constitutes hypertension, may be subject to recall bias if asked in an interview or questionnaire or, if the study is record-based, to differential availability of information. Obtaining an actual blood pressure reading eliminates all of these problems, but there may still remain errors due to variability of the measurement or to judgment on the part of the observer. In this instance, the most valid estimate would result from taking an average of several blood pressure readings, with trained observers following a detailed, standardized protocol [8, 9].

It is important to remember that for epidemiologic purposes, the less room there is for interpretation by either the investigator or the subject, the less likely it is that bias will occur. While the question "How do you feel?" may be clinically relevant in terms of eliciting information helpful to making a diagnosis, a more useful question epidemiologically would be "Have you experienced any of the following symptoms?" followed by a comprehensive list of specific conditions or reactions of interest.

With respect to the administration of the data collection instrument, the single most important way to minimize the potential for bias is to maintain blindness to the greatest extent possible. In practical terms, this means that study personnel who abstract records and interview or examine subjects should be unaware of an individual's exposure status when determining the outcome of interest in a cohort or intervention study, or an individual's disease status when assessing exposure in a case-control study. Furthermore, subjects should be kept as unaware of their own study status as well as of the specific hypotheses under investigation as is logistically feasible and ethical. For example, in a case-control study of alcohol consumption in relation to MI [2], the way in which the study was introduced to potential participants was as an investigation of risk factors for serious illness requiring hospitalization. Consequently, while the subjects certainly knew they had been hospitalized for MI, they were not aware that it was the diagnosis of MI rather than hospitalization per se that was the selection criterion used. Moreover, the investigators obtained information on a wide variety of factors, including demographic variables, medical history, family history, physical activity, personality type, medication use, smoking, coffee drinking, and an extensive dietary assessment, so that alcohol consumption was only one among dozens of variables mentioned. In these ways, the major hypothesis under investigation was masked to minimize the potential for bias in the recall of information by the subjects or the collection of the information by the interviewers.

While the effectiveness of blindness as a means to minimize bias is well recognized, it is not always possible to achieve. Studies of well-publicized hypotheses will often be difficult to disguise, and participants who are affected or unaffected by either the exposure or the outcome may respond differently according to their understanding of the aims of the study and their belief in the existence of the hypothesized association. Similarly, in case-control studies of serious illnesses, especially those that utilize community controls, it is inevitable that interviewers will know a participant's status as a case or control. Clinical trials testing interventions for which placebo control is either impossible or undesirable will also be unable to maintain blindness. When it is not possible to incorporate blinding of the subjects and/or investigators into the study design, it is especially important that other approaches known to minimize observation bias be utilized. For example, if those administering the treatment in a clinical trial that is not placebo-controlled are aware of the subject's exposure status, it would be most important for different study personnel who are blinded to exposure status to ascertain the endpoints of interest.

A second means to minimize the potential for bias in the administration of data collection instruments is the implementation of rigorous, standardized training of study personnel and the use of clearly written protocols. To decrease the likelihood of observation bias, it is important that any individuals who complete forms, examine subjects, or administer questionnaires follow very specific procedures that are identical for all subjects. Training in such procedures should include standardizing responses to questions about the study, adopting uniform ways of probing for additional information, and employing standard techniques for dealing with errors or missing information.

As a general check for observer bias, a number of items can be included in the data collection instrument that may alert the investigator to the presence of problems. First, study subjects may be compared with respect to their frequency of reporting (or other ascertainment) of dummy variables, which are factors believed to be either unrelated to the exposure or disease of interest or clearly related but in a known way. For example, in a case-control study of regular aspirin use in relation to risk of MI, information could be collected on a wide range of pharmacologic exposures, such as other analgesics, that are not known or sus-

pected risk factors for MI. If cases and controls differ with respect to their reported use of aspirin but not in their use of other analgesics, these data would support the belief that the observed difference in reported aspirin use is real rather than a function of the study design or conduct. On the other hand, if reported use of both aspirin and other analgesics is higher in the cases than in the controls, the suspicion of some type of recall or interviewer bias must be raised. Similarly, information on a number of factors known to increase risk of MI could be obtained, such as history of hypertension or family history of coronary heart disease. The ability to reproduce these known associations using the study data would again add support to the belief in the validity of any findings regarding aspirin and MI. Second, it is possible to construct a questionnaire or interview that includes several items seeking the same information in different ways, in order to ascertain systematic differences in probing for specific responses. Third, the characteristics of the data collection procedures themselves can be examined, including recording the time an interview, examination, or form completion began and ended in order to check that study personnel are not systematically spending more or less time probing for information depending on an individual's exposure or disease status. Finally, it may be useful for the examiner to record a simple reliability score that reflects the interviewer's subjective assessment as to the participant's ability to understand the questions posed and respond reliably. Study subjects whose information is deemed unreliable may then be excluded from the study or analyzed separately.

Sources of Exposure and Disease Information

In addition to the means of collecting data, the number and nature of sources of exposure and disease information in a particular study will also affect the potential for bias. Information can be obtained from sources as varied as questioning the participants themselves, utilizing hospital or employment records or vital statistics data, or measuring variables directly. The use of preexisting records is the most unbiased source for such data, since the information was recorded prior to the onset of an outcome event. However, preexisting records may not have complete information on all factors of interest for each study subject, especially with respect to lifestyle variables such as smoking, exercise, or diet. Moreover, the amount of missing information may vary for different study groups. For example, in a case-control study of conjugated estrogen use in relation to uterine cancer [1], hospital records were used to identify exposure among the cases as well as a comparable group of women admitted for nonmalignant conditions. For each study participant, information on medication use was also obtained from her private

gynecologist's records. With a positive notation of estrogen use from either the hospital or physician's records taken to indicate "true" use, 42 percent of current users among the controls would have been classified as nonusers on the basis of hospital charts alone, as compared with only 15 percent of women with uterine cancer. The most likely explanation for this differential recording of exposure information in these preexisting records is that knowledge or suspicion of a possible relationship between estrogen use and cancer caused staff members at the hospital to probe more thoroughly for details of drug use among women with cancer than among those with nonmalignant conditions.

As this example illustrates, one way to minimize this potential for bias is to use multiple sources of data whenever possible as a way to provide an independent verification of exposure or disease status. Record-based studies could use both hospital and physicians' records, for example, while interview or questionnaire data could be confirmed or supplemented through examination of medical records. Self-reported risk factors and diagnoses are frequently documented through examination of the relevant hospital discharge summaries, pathology reports, or other medical records. Diagnoses reported on a death certificate might be confirmed by abstracting information from hospital records if the subject died after admission, or by examining additional details about the death obtained from relatives. Similarly, in case-control studies, the identification of cases from hospital discharge summaries or pathology reports could be confirmed through a standardized, independent review of records and slides by a single study pathologist blinded to the exposure status of the subjects. In intervention studies, there is often an attempt to corroborate self-reports of compliance with the study regimen through the examination of biochemical or other objective markers of adherence. In all these instances, the goal is to provide evidence of exposure or disease status that is complete and unlikely to be subject to bias on the part of the study participants or observers.

The need for supplemental information will depend to a great extent on the nature of the particular outcome of interest [3]. All exposures and outcomes of interest should be carefully defined using standard, uniform criteria to minimize the need for interpretation on the part of study personnel. For example, many investigations of MI utilize criteria established by the World Health Organization [14]. It is also important that the individuals responsible for assigning such diagnoses be unaware of the subject's exposure status, since it is often difficult to eliminate completely the need for judgments on the part of the reviewer.

Finally, during the planning stages of any investigation, practical methods for minimizing losses to follow-up and obtaining complete ascertainment of outcome should be incorporated into the study design. Depending on the hypotheses being tested and the specific design features of the study, such methods might include, at a minimum, use of

the National Death Index established by the National Center for Health Statistics [13], systematic searches of state vital statistics records, telephone calls to nonrespondents or dropouts, and brief written communications designed to ascertain vital status. For such strategies to minimize bias effectively, it is essential that they be implemented according to standardized protocols that apply equally to all participants, regardless of exposure or disease status.

EVALUATION OF THE ROLE OF BIAS

To the fullest extent possible, potential sources of bias should be eliminated or minimized through rigorous design and meticulous conduct of a study, as discussed in previous sections of this chapter. In practice, however, it is rarely possible to know with assurance that such efforts have been successful. Consequently, for any individual study it is always necessary to consider carefully what possible biases may have influenced the observed results; the direction of the likely effect, that is, whether the bias would have acted to mask a true association or to cause a spurious association where there truly is none; and, if possible, how great this distortion might be. This will involve consideration of the type of analytic study utilized (i.e., case-control, cohort, or intervention), specific aspects of the study design and conduct, and the nature of the findings (i.e., whether the results are reported as "null" or representing the presence of an association). Often when a study is published, a general criticism is raised that inaccuracies in the ascertainment of exposure or disease or the nature of the study participants themselves cast serious doubt on the validity of the findings. This criticism may be voiced despite the fact that sometimes the action of the putative bias is not even in a direction that could have accounted for the suggested result. For example, it would be inappropriate to suggest that the previously cited observed increased risk of congenital malformation among women who were identified as having filled a prescription for spermicides within 600 days before birth [7] is due to misclassification of exposure. While there is no doubt that such misclassification occurred and may even have been substantial, there is no reason to believe that the inaccuracies were differential between cases and controls, since the exposure information was recorded before the outcome was known. Such random misclassification could only have led to an observed underestimate of the effect of spermicide use on risk of congenital malformation. It could never have caused the observation of a spurious association.

In evaluating the likelihood that a potential bias may have affected the study results, investigators can sometimes take advantage of evidence from some aspect of the study itself. For example, in their early case-control study of smoking and lung cancer, Doll and Hill [4] found a higher frequency of those reporting heavy smoking among the cases than the controls. One concern in interpreting these findings was that the interviewers probed more deeply for a history of smoking among those with lung cancer than controls without the disease. Evidence to evaluate this possibility was provided by the fact that during the course of the same investigation, a number of patients were interviewed in the belief that they had lung cancer but were subsequently determined not to have this diagnosis. If observation bias were responsible for the increased proportion of smokers observed among the cases, then the subjects who were incorrectly diagnosed as having lung cancer should have had smoking histories similar to those of the actual lung cancer cases. In fact, the smoking histories of the incorrectly diagnosed patients more closely resembled those of the controls, suggesting that the observed difference in smoking habits was not the result of bias on the part of the patient or interviewer. Similarly, in the evaluation of the association between postmenopausal hormones and uterine cancer, the possibility was raised that subclinical cancers were being diagnosed more frequently in women using these agents as a result of closer medical surveillance and diagnostic evaluations that otherwise would not have been performed [5]. The fact that the association persisted even when such early in situ cases were excluded [6] supported the belief that such selection bias, although still a potential problem, could not have accounted for the entire increased risk observed.

In other circumstances, investigators may choose deliberately to build into the study design certain strategies to allow them to assess the extent to which a particular potential bias actually affects either their own or previously reported findings. For example, to address the issue of recall bias in studies of spermicide use and congenital malformation, a case-control study was conducted in which reported use of these preparations by mothers of children with Down's syndrome was compared with that of mothers of both healthy infants and children with congenital heart disease, a condition that in previous studies had not been associated with spermicide use [10]. This second control series was chosen specifically because the children were also ill and thus presumably the information provided by their mothers would be subject to the same recall bias as that from mothers of cases. In this manner, it was possible to evaluate the role of recall bias as an alternative explanation for the observed association between spermicides and Down's syndrome. Although the magnitude of the effect was somewhat smaller than the result using the nondiseased controls, the association between Down's syndrome and spermicide use persisted even when this second control series was examined.

In most circumstances, an evaluation of the role of bias as an alternative explanation of the study findings must involve considerations as to the type of study, its particular design, and the nature of the results.

While all analytic studies are subject to bias, each design has particular types to which it is inherently most vulnerable. For example, if a study uses the case-control approach, concern about the potential for bias would focus primarily upon two possibilities: (1) that knowledge of disease status influenced the determination of exposure status (recall bias), and (2) that knowledge of exposure status influenced the identification of diseased and nondiseased study subjects (selection bias). In a cohort study, the possibility for bias due to loss to follow-up would be a particular concern, as would selection bias if the study were retrospective. On the other hand, if the study were a prospective cohort investigation, selection bias would not be a major concern. In an intervention study, the degree of concern about the potential for observation bias would depend on the nature of the comparison group, the use of a placebo, and the degree of objectivity in the measurement of the outcome.

The potential for misclassification must also be considered in every type of epidemiologic study. The most important issue in this regard is whether the inaccuracies in classification of exposure or disease are differential or random. If differential, the misclassification could result in either an underestimate or overestimate of effect, depending on the direction of the error; for example, in a case-control study, whether the cases are more or less likely to report a positive exposure history than the controls. On the other hand, many inaccuracies in ascertainment of exposure or disease are inevitable in all epidemiologic research and are thus present to a similar degree in every study design. If the misclassification is likely to be random, that is, if there is no reason to believe that the level of error would be different in the different study groups, then such bias can only work in the direction of underestimating the study results. Therefore, in interpreting a particular study, if the results indicate the presence of an effect it is not possible to account for this result by any amount of random misclassification, no matter how substantial, since any estimate of the effect without the misclassification could only by more extreme. On the other hand, random misclassification could certainly be responsible, totally or in part, for a finding of no association between exposure and disease, depending on the size of the true effect and the magnitude of the misclassification, and must always be considered as a possible alternative explanation for a reported null finding.

CONCLUSION

Bias, like chance and confounding, should always be considered as a possible alternative explanation of any observed statistical association, whether positive, inverse, or null. Unlike chance and confounding, however, which can generally be taken into account in the analysis through the use of appropriate statistical techniques, bias is most effectively dealt with through careful design and meticulous conduct of a study. Once a potential source of bias is introduced, it is usually extremely difficult to correct for its effects analytically. It is, however, possible as well as necessary to identify likely sources of bias in a particular study and attempt to estimate both the direction in which the bias would have altered the estimate of effect and, when feasible, the magnitude of that influence. Investigators should discuss all these issues fully in published reports to provide readers with the maximum opportunity to judge for themselves whether bias is likely to account for observed findings. However, whether or not investigators follow this guideline, readers should always, to the extent possible, consider the potential role of bias as an explanation for reported results.

STUDY QUESTIONS

1. Compare and contrast the likelihood of occurrence of selection and observation bias in case-control and cohort studies.
2. Compare and contrast the effects of random and nonrandom misclassification in the estimate of relative risk.
3. What are the major options in the design of a study to minimize the occurrence of bias?

REFERENCES

1. Buring, J. E., Bain, C. J., and Hennekens, C. H. Alternative data sources in a case-control study of conjugated estrogens and cancer. *Am. J. Prev. Med.* 2:116–121, 1986.
2. Buring, J. E., Willett, W., Goldhaber, S. Z., et al. and the Boston Area Health Study Research Group. Alcohol and HDL in non-fatal myocardial infarction: Preliminary results from a case-control study. *Circulation* 68:227, 1983.
3. Colditz, G. A., Martin, P., Stampfer, J. J., et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of disease in women. *Am. J. Epidemiol.* 123:894–900, 1986.
4. Doll, R., and Hill, A. B. Smoking and carcinoma of the lung: Preliminary report. *Br. Med. J.* 2:739–748, 1950.
5. Feinstein, A. R., and Horwitz, R. I. A critique of the statistical evidence associating estrogens with endometrial cancer. *Cancer Research* 38:4001–4005, 1978.
6. Gordon, J., Reagan, J. W., Finkle, W. D., et al. Estrogen and endometrial carcinoma. An independent pathology review supporting original risk estimate. *N. Engl. J. Med.* 297:570–571, 1977.
7. Jick, H., Walker, A. M., Rothman, K. J., et al. Vaginal spermicides and congenital disorders. *J.A.M.A.* 245:1329–1332, 1981.

8. Rosner, B., and Polk, B. F. Predictive values of routine blood pressure measurements in screening for hypertension. *Am. J. Epidemiol.* 117:429–442, 1983.
9. Rosner, B., Cook, N. R., Evans, D. A., et al. Reproducibility and predictive values of routine blood pressure measurements in children: Comparison with adult values and implications for screening children for elevated blood pressure. *Am. J. Epidemiol.* In press, 1987.
10. Rothman, K. J. Spermicide use and Down's syndrome. *Am. J. Public Health* 72:399–401, 1982.
11. Sackett, D. L. Bias in analytic research. *J. Chronic Dis.* 32:51–63, 1979.
12. Sartwell, P. E., Masi, A. T., Arthes, F. G., et al. Thromboembolism and oral contraceptives: An epidemiologic case-control study. *Am. J. Epidemiol.* 90:365–380, 1969.
13. Stampfer, M. J., Willett, W. C., Speizer, F. E., et al. Test of the National Death Index. *Am. J. Epidemiol.* 119:837–893, 1984.
14. World Health Organization. *IHD Registers: Report of the Fifth Working Group.* Copenhagen: World Health Organization, 1971.

12

Analysis of Epidemiologic Studies: Evaluating the Role of Confounding

In Chapters 10 and 11, we discussed methods for evaluating the effects of two possible alternative explanations that must always be considered in assessing the presence of a valid statistical association in a given study—the role of chance and that of bias. The third alternative explanation, confounding, involves the possibility that the observed association is due, totally or in part, to the effects of differences between the study groups other than the exposure under study that could affect their risk of developing the outcome of interest. The concept of confounding is central to the interpretation of the findings of any epidemiologic study—most critically in observational studies but also for experimental investigations. Unlike bias, which is primarily introduced by the investigator or study participants, confounding is a function of the complex interrelationships between various exposures and disease.

THE NATURE OF CONFOUNDING

Intuitively, confounding can be thought of as a mixing of the effect of the exposure under study on the disease with that of a third factor. This third factor must be associated with the exposure and, independent of that exposure, be a risk factor for the disease. In such circumstances the observed relationship between the exposure and disease can be attributable, totally or in part, to the effect of the confounder. Confounding can lead to an overestimate or underestimate of the true association between exposure and disease and can even change the direction of the observed effect. For example, consider a study that showed a relationship between increased level of physical activity and decreased risk of myocardial infarction (MI). One additional variable that might affect the observed magnitude of this association is age. People who exercise heavily tend to be younger, as a group, than those who do not exercise. Moreover, independent of exercise, younger individuals have a lower risk of MI than older people. Thus, those who exercise could have a lower risk