Chapter 7

Case-Control Studies

Case-control studies are closely related to prevalence or cross-sectional studies (discussed in Chap. 6). However, because they generally involve fewer and more readily accessible subjects, case-control studies are much more often carried out. Among analytic studies, they are usually the first approach to determining whether a particular personal characteristic or environmental factor is related to disease occurrence.

How Case-Control Studies Are Carried Out

Identification and Collection of Cases Once the study objectives and methods have been clearly defined, the first step in a case-control study is the identification of the cases or diseased persons to be studied. (Many rightfully object to the use of the term "case" to refer to a sick human being. Although this dehumanizing

term should be avoided in the clinical setting, its use facilitates clear communication about research. In this context it does not imply any lack of sympathy or concern about the ill.)

As mentioned previously in connection with prevalence studies, it is important to set up criteria for the diagnosis and inclusion of cases in the investigation and to describe these criteria carefully when the study is finally reported. It is usually advisable to require objective evidence and documentation of the disease, even if, as a result, some cases will have to be omitted and the size of the case group reduced. Thus, for a study of renal calculi, it may be best to insist that all included cases have stones documented by x-ray evidence or removal by surgery, not diagnosed only by the presence of renal colic. By accepting less well-documented cases, the investigator runs the risk of diluting his case group with some noncases and lessening his chances of finding differences between the case group and the control group.

This recommendation, of course, applies to disease identification for all types of studies, not just case-control studies. However, as was stressed in the last section of Chap. 3, *misclassification* of a few nondiseased persons as cases and of a few diseased persons as controls, no matter how distressing to the clinician, will probably not prevent the discovery of major case-control differences.

The cases may be identified or "ascertained" by a community-wide search, but more often, they are limited to those found in one, or perhaps a few, hospitals, clinics, or medical centers. The case group will usually be limited to those seen or diagnosed during a particular time period. For example, one may decide to study all cases of well-documented renal stones seen at a particular hospital during the 2-year period, January 1, 1974 through December 31, 1975.

Usually, it will not be possible to include in the study all the patients who meet the diagnostic criteria and the time and place specifications. There will be a variety of reasons for this. Some patients will have moved away, died, or will refuse to cooperate; or, some hospital records may be lost so that certain essential information is not available to the investigator. He or she, in turn, should report how many cases met the initial criteria for inclusion and how many were finally included. The reasons why some cases had to be

omitted and the number of cases omitted for each reason should be stated.

Selection of Control Subjects The decision as to who will constitute the control group or groups is perhaps the most difficult one to be made in planning a case-control study, and it requires a good deal of skill and judgment. In a prevalence study this problem does not arise since the cases may be compared with the entire nonaffected portion of the population. By settling for the simple low-cost case-control study instead of the large community-wide prevalence study, the investigator gives up the chance of comparing all the diseased and nondiseased persons in the community. However this is done in the hope that almost as much can be learned about the relationship of the disease to other variables by studying a group of cases and a group of controls. Sometimes a relatively small sample derived randomly from the entire population can be utilized as a control group. However obtaining the desired participation of this kind of representative control group is difficult and often not feasible.

General Principles One of the most important considerations in selecting controls involves the information to be collected concerning study variables or potential etiologic factors. There should be no major differences between case and control groups as to the quality or availability of this information. Availability of information implies both (1) how much information is obtained concerning each case and control, and (2) what proportions of the case and control groups will, or can, supply it. Equal access to important information previously recorded in a similar fashion for both cases and controls—for example, birth weight recorded in the same hospital—may strongly favor the use of a particular control group. If data have to be obtained by interview, then one worries that quality or availability of information may differ due to differences between cases and controls in emotional state, knowledge of the disease studied, educational or socioeconomic status, and location of the interview (e.g., at home or in a hospital).

Consideration of the *known* sources of bias in quality and quantity of information about cases and controls and of the fact that there are often biases which are *unknown* usually leads the investigator to attempt to find controls that are similar in a general way to the cases, except for the essential difference in whether the disease under investigation is present or absent. Yet, this striving for general similarity should not be carried to the point where there is little or no hope of finding case-control differences in the factors under study. For example, by selecting the controls so that they are of similar educational background to the cases, one will minimize case-control differences in the understanding of a written questionnaire. But this selection procedure will also preclude the study of the relation of educational level to the disease and may seriously impair case-control comparisons of factors related to education, such as socioeconomic status.

In selecting a control group two major questions must be answered

1 From what source(s) will controls be drawn?
2 What will be the method of selection of controls from each of these sources?

These decisions must take into account the need, mentioned above, for controls that are generally similar, but not too similar, to the cases, plus some very practical considerations—in particular, the control groups that are potentially available, and the human and financial resources that can be used for the study.

Selecting a Source of Controls Many sources of controls have been used, including:

1 Patients within the same medical-care facility
   a Without regard to their diagnosis
   b Excluding those with certain diseases
   c Including only those with certain diseases such as mild or "act-of-God" conditions (e.g., hernias, accidental injuries)
   d Examined and found to be healthy
2 Persons drawn from outside the facility
   a Sample of general community
   b Friends or acquaintances
   c Fellow employees
   d Neighbors
   e Family members such as spouses or siblings

When one is faced with the practical decision as to which source of controls to use, reasons for and against any potential source can usually be mustered, and the reasons why the source chosen might have given biased results will be heard from critics after the study is reported. For example, the investigator may decide to select controls for hospitalized renal calculus cases from herniorrhaphy cases in the same hospital, since that hospital serves a particular socioeconomic and ethnic segment of the community, and since, after the acute pain has subsided, the mental status of a kidney stone patient should not be very different from that of a hernia patient (as contrasted with a patient, say, with a stroke or terminal cancer). Yet if an important difference between kidney stone patients and their hernia controls is found, there will usually be the lingering question of whether the difference is related to kidney stones or to hernias. Therefore, it is frequently helpful to have a diagnostically heterogeneous control group, or more than one control group, if possible. Similarly, repetition of the study by other investigators in other settings will usually reveal whether or not some underlying truth about renal calculi has been discovered. MacMahon and Pugh (1970) have thoroughly discussed many of these important issues and other factors to be considered in selecting controls.

Selecting Control Subjects from the Source Selection of the control group from the chosen source usually involves sampling. If resources are limited, the control group will usually be equal in size to the case group or smaller than the case group, if necessary. If resources permit the inclusion of more study subjects and no more cases are available, the control group may be enlarged to decrease sampling variation by having, for example, twice or three times as many controls as cases, or even more.

As already noted, selecting a source places some general limitations on the nature of the control group. In addition, when individual controls are chosen from the source, the investigator will often *match* the controls to the cases with regard to some important characteristics such as age or sex. By matching on a particular characteristic, the investigator immediately eliminates a case-control difference in this characteristic as a possible contributor to a case-control difference in a study variable. For example, if the cases

and controls are matched for age and it is subsequently found that they differ in blood pressure, age could not be the explanation for this blood pressure difference. In the unusual instance that nothing is known about the disease, not even, say, its age and sex distribution, then no matching would be desired since matching precludes any case-control comparison of the matched variable.

Controls are usually picked individually, in a "paired" fashion. That is, for each case, one or more controls is picked in some systematic fashion according to preset rules or criteria. In a study of renal calculi, it may be decided to include as controls other urological patients who have no urinary-tract stones or obvious mental impairment due to uremia or other cause and who are matched to the cases with regard to age, sex, race, and date of admission. The paired selection of a matched control for each case might involve selecting the first patient admitted to the urological service after the case, who meets the diagnostic and mental status criteria, who is of the same sex and race as the case, and whose age differs by no more than 5 years from that of the case. Some leeway is necessary in matching for quantitative variables such as age and admission date, or else no match will be found for most cases. Failure to find matched controls will also occur frequently if matching is attempted on more than a few characteristics.

If the disease being studied is known to be uncommon in the group serving as a source for controls, then little, if any, diagnostic effort or documentation is needed to rule out the disease in the selected controls. However, if the disease could occur commonly in controls, at least some attempt to rule it out, such as an interview question or a quick review of the medical chart, is desirable to minimize misclassification.

Data Collection Any source of data about the study variables may be used. As has been mentioned, accurate information collected on both cases and controls before the disease developed is ideal. Collecting information after the disease develops may be necessary, but every effort should be made to avoid qualitative and quantitative case-control differences in the data gathered. For example, if possible, the research assistant(s) recording laboratory data for all study subjects should do so without knowing whether

particular individuals are cases or controls. Similarly it may often be desirable to structure data-collecting interviews to avoid discussing disease status altogether, or at least until the questions about etiologic variables have been asked.

Data Analysis Normally, the basic case-control comparison is expressed in terms of the proportion of cases versus the proportion of controls who show a particular characteristic. If the characteristic is quantitative rather than a qualitative "yes-or-no" attribute, then its distribution in cases and controls can be compared, as can the more general descriptions of the distribution, such as the mean, standard deviation, and the median.

Interpretation

If the cases show a higher proportion with an attribute than do the controls or if the distributions or mean levels of an attribute differ, then there is an observed association between the attribute and the disease. Interpreting whether this association implies a cause-and-effect relationship is another matter, involving a number of considerations to be discussed in Chap. 11.

It may seem more convenient or natural to think about the study results expressed, as is usually done in a prevalence or incidence study, as the rate of disease occurrence in persons with a particular attribute compared to the disease rate in those either without that attribute or with a different attribute. In case-control studies the results of comparisons are usually expressed in the converse manner, that is, as the relative frequency of the attribute in the diseased versus the nondiseased. Fortunately, the results of case-control studies can be converted mathematically to comparisons of disease rates, or at least to an expression of relative risk of disease, under certain conditions. These are, that cases and controls are reasonably representative of persons with and without disease in the underlying population and that the disease prevalence rate of the underlying population is known, or at least known to be small. The interested reader should refer to MacMahon and Pugh (1970) for a description of these methods.

As with prevalence studies, case-control studies usually involve

existing disease cases which, as discussed in Chap. 6, p. 80, may differ in a variety of ways from all cases that develop. One way to try to overcome this problem is to include only those cases that first develop or are first diagnosed during the period of data collection. By using only new cases and selecting controls to be representative of the population at risk for developing the disease, the case-control study then aims more directly at determining factors responsible for disease *development*, much like an incidence study. Paradoxically, although this should provide a broader and more representative *spectrum* of cases, it may limit the *number* of cases available for study, resulting in a sample size that is too small to provide reliable data.

It should also be emphasized that the source of cases for the study may be more apt to provide medical care to one type of case than another. For example, cases derived only from a hospital and not from outpatient clinics as well, may have the most severe disease. Thus, while we have emphasized the problems and vagaries of control groups, the characteristics of the case group must also be carefully considered in study design and interpretation.

Example 1: Oral Contraceptives and Thromboembolic Disease

Millions of women now take oral contraceptive tablets to prevent pregnancy. Several questions concerning the safety of these agents have arisen. One of the major areas of concern has been whether or not oral contraceptives predispose to thromboembolic conditions, particularly thrombophlebitis and its possibly fatal sequela, pulmonary embolism. Following the publication of some clinical case reports in the early 1960's it became apparent that epidemiologic studies were necessary to determine whether women who take oral contraceptives are indeed at greater risk of developing these diseases.

Thrombophlebitis and pulmonary embolism *not* secondary to trauma, surgery, or childbirth, develop rather rarely in women during the reproductive years. Thus a prevalence or incidence study of this question seemed impractical, at least as a first approach, since many thousands of women would have to have been studied in order

to find an adequate sample of cases. Case-control investigations were therefore undertaken, both in Great Britain and the United States. The U.S. study by Sartwell and his associates is an excellent example of the case-control method.

The investigators decided to include as cases, women, ages 15–44, hospitalized with thromboembolic conditions and discharged alive within the previous 3 years. It was necessary to collect the cases from a large number of hospitals to obtain an adequate sample size. All told, there were 48 participating hospitals in five large eastern cities: Baltimore, New York City, Philadelphia, Pittsburgh, and Washington, D.C. Cases were excluded from the study if they also had a chronic condition possibly predisposing to thromboembolism, such as diabetes mellitus or hypertension, or a recent precipitating event such as surgery, pregnancy, trauma, localized infection, or prolonged inactivity. Reasonable medical evidence for thromboembolism was required, and all cases were reviewed independently by two physicians.

The derivation of the final study group of 175 cases was carefully described by the authors and clearly shows the marked attrition that often occurs between *potential* and *actual* numbers of study subjects. In all, 2,648 women in the desired age range with thromboembolic conditions within 3 years were identified and their hospital records were abstracted. The vast majority of these cases, 2,288, were immediately rejected because of having possibly predisposing conditions, and another 99 were rejected for other reasons, such as sterility (which obviates contraceptive use), death, or having moved from the area. Of the 261 women selected as suitable cases, 72 had to be dropped because the interview could not be obtained and another 14 were excluded because no interview could be obtained from their matched control subjects.

Two matched controls were selected for each case with the expectation that if one could not be interviewed the alternate control would still be available, thus yielding data on one control per case. Matching was done on several criteria:

Hospital	:	same as case
Sex	:	all women
Discharge date	:	same 6-month interval as that of case

Discharge status	:	all alive
Age	:	same 5-year span
Marital status	:	same
Residence	:	(not stated but presumably the same metropolitan area)
Race	:	same
Parity	:	same general class, i.e., no pregnancies, one or two pregnancies, three or more pregnancies
Hospital pay status	:	ward, semiprivate, or private room

Also, controls were excluded in the same manner as the cases, i.e., for chronic diseases possibly predisposing to thromboembolism or for sterility. Most control subjects turned out to have acute medical and surgical illnesses, conditions treated by elective surgery, or traumatic injuries.

Cases and controls were interviewed at home. A variety of questions were asked so as to provide data concerning pertinent variables such as religion, educational level, and smoking habits. To elicit information about contraceptive usage, cases and controls were asked to select from a list of thirteen methods those which they had used within the 2 years before they were hospitalized.

Data analysis showed that the overall frequency of employment of *any* birth-control method was similar in the 175 cases and controls—114 and 101 users of at least one method, respectively—and many women had used more than one method during the 2-year period. While the case-control differences in proportions using each of the other methods were small and not statistically significant, cases did report using oral contraceptives significantly more often than did controls—67 versus 30 women or 38 percent versus 17 percent.

Using a simple formula to compute relative risk, the investigators found that users of oral contraceptives were about four times as likely as nonusers to develop thromboembolic conditions. Furthermore it could be shown that about one-fourth of the total cases would be attributable to oral contraceptive usage if a cause-and-effect relationship were involved. It was, of course, carefully pointed out that the cases studied were a highly selected group, that is, free

of predisposing conditions, unlike most thromboembolism cases.

Further analysis showed that the case-control differences in oral-contraceptive use were present in the major subgroups of the study subjects, when the total group was subdivided by such variables as age and marital status. The case-control differences were found for several different thromboembolic conditions including deep thrombophlebitis of the lower extremity, pulmonary embolism, and intracranial vascular conditions.

Example 2: Pedestrians Fatally Injured by Motor Vehicles

In their concern with learning about the diseases which present complex diagnostic or pathophysiologic problems, medical personnel are apt to forget that injuries and death due to gross physical trauma are one of the chief health problems in affluent industrialized societies as well as in "less developed" areas. In particular, accidents are the leading cause of death in children and young adults in the United States. Automobile accidents lead all other types as a cause of death.

The word "accident" implies that physical injuries produced by automobiles and other energy sources are haphazard and uncontrollable. Among those arguing against this fatalistic concept, Haddon has advocated the use of carefully designed and implemented epidemiologic studies as a means of identifying factors responsible for traumatic injuries, so that appropriate preventive measures can be instituted. His research group's interesting study of the characteristics of pedestrians fatally injuried by motor vehicles in New York City is an example of the imaginative use of the case-control method to attack a serious and poorly understood problem (Haddon et al., 1961).

At the time of the study in 1959, little was known about pedestrian-associated or "host" factors related to being struck and killed by a car. Substantial funds were being expended for public education programs and other means of "pedestrian control," without much evidence that these were effective preventive measures. The previous findings that many fatally injured pedestrians had been drinking heavily had not been evaluated in comparison to

the alcohol consumption of the population at risk or, more simply, to that of noninjured pedestrians. Likewise, the age distribution of killed pedestrians, with relatively high percentages of young children and elderly adults, had not been compared with the age distribution of all or of nonkilled pedestrians, to determine whether the *mortality rate*, or risk of being killed, is actually greater in very young and very old pedestrians. Thus, age and blood-alcohol concentration were included among several characteristics that were measured in fatally injured pedestrians and their matched controls in the study to be described.

New York City was a very appropriate place for this investigation. Pedestrian deaths were relatively frequent, and they accounted for about 70 percent of all fatalities in motor vehicle accidents. The case series consisted of 50 adults (18 years of age and older) who were struck and killed by automobiles in Manhattan between May 3, 1959 and November 7, 1959. Autopsy confirmation of the cause of death was required. Of 57 cases initially considered, the 7 omissions consisted of 2 who were killed by bicycles, 1 who was purposely pushed into the path of a car, 1 with unknown site or time of the accident, 1 who died of a coronary occlusion while convalescing from the accident, and 2 who were omitted because of clerical errors.

Four matched controls were selected for each case by visiting each accident site at a later date, but on the same day of the week and as close as possible to the time of day when the accident occurred. All but eight site visits for control selection were completed within 6 weeks of the accident. Thus, controls were matched to the cases for accident site and time. In addition, controls were matched to the accident cases for sex and were limited, as were the cases, to adults.

The practical problems involved in this form of "shoe-leather" epidemiology can best be communicated by the investigators' own description of the control selection and interview procedures:

> The site visits were made by a team of two or three of the authors and one to four medical students working at each location with one or two uniformed members of the Police Department Accident Investigation Squad (A.I.S.).

In visiting each site one of three basic approaches was used. In the first type, that used in many busy neighborhoods, for example, opposite Grand Central Station on a weekday at 6:10 P.M., the entire team arrived and immediately stopped the *first* 4 adult pedestrians of the same sex as the deceased. At such busy sites the group arrived and accomplished its purposes in 5 minutes or less from start to finish.

When the accident site was in a neighborhood in which it was suspected that the group might be seen and avoided, a second approach was used. Under such circumstances, for example, at sites in the Bowery, the group arrived and 'swept the block' stopping successively the *first* 4 adult pedestrians of the required sex who were headed toward or away from the accident site. By pedestrian here and throughout this report is meant a person progressing by walking, not lounging stationary, sitting, or lying down.

In the third approach, used where pedestrian traffic was very light, for example at 108th Street and the East River (F.D.R.) Drive at 1:40 A.M., the group would lounge nearby or sit in a car at or near the site watching for approaching pedestrians, and as each of the *first* 4 of these came into view he, or, where appropriate, she, was quickly approached and stopped.

The site visited was the sidewalk point closest to the exact location of the accident as described on the police or medical examiner's report. For example, one report indicated that the deceased had been crossing the street 40 feet from a given corner. This was found to be directly in front of a 'rathskeller', and it was at that point that the first 4 pedestrians were stopped.

Great care was taken to avoid any attempt at matching for the characteristics of the deceased, except in so far as sex and adulthood were concerned. In addition, for methodologic uniformity, at all sites the same investigator pointed out to the accompanying police each individual to be stopped. Although the exact details varied with the circumstances, the person was immediately approached and told by the policeman, 'Please step over for a minute while the doctors ask you a few questions.' A nearby member of the team immediately stepped up and began talking uninterruptedly: 'I don't want to know your name; I merely want to ask you a few questions. Do you live in Manhattan?' The interview was usually easily begun in this manner, although 12 refusals occurred (for each of which the next pedestrian was substituted) . . . .

This investigation was carried out without publicity of any kind. With one exception it was invariably possible to stop the members of each pedestrian sample prior to the formation of the substantial group of watchers which sometimes formed thereafter. The exception, in a 'tough' neighborhood at 2:30 A.M., involved the only site at which 2 persons had been fatally injured in the same accident. On arrival, it was possible to obtain quickly the first 7 but not the eighth interview and specimen of breath, a small, hostile crowd quickly forming from an adjacent bar. As a result, only the first 4 of the 7 interviews and specimens obtained at this site were used, being counted twice in the analyses of the data.

The interview included questions as to: place and length of residence; place of birth; age; present occupation; and marital status. Sex, apparent race, appearance and apparent sobriety, date, location, time of interview, and weather were also recorded.

Immediately on finishing the interview the interviewer stated approximately as follows, 'I only have one more thing for you to do (and then you can go) and that is to blow up this bag for me.' Simultaneously he removed a Saran bag from an envelope and showed the pedestrian how to place one of its two ends in his mouth and blow until told to stop. This finished, the pedestrian was thanked and told that the interview was over.

A large percentage of those interviewed were foreign born, and many of these admitted to no knowledge of English. Rather than weaken the investigation by omitting these pedestrians when no member of the team knew a common language, passersby were stopped and asked to serve as interpreters. Apparently because those walking in the same neighborhoods or, in some cases, accompanying those stopped (many of the latter being interviewed themselves) tended to know the same languages, this procedure proved very satisfactory. With its use no one failed to be interviewed because of a language barrier and interviews were completed in Armenian, German, Greek, Spanish, and other languages and dialects.

As implied above, blood-alcohol concentrations were measured by analysis of breath specimens and the other data concerning the controls were recorded as described. Data concerning the cases were obtained chiefly from official records describing the accidents.

Postmortem blood-alcohol measurements were studied in those cases who survived less than 6 hours after the accident.

Data analysis for the case-control comparison revealed that, indeed, fatally injured pedestrians were older than the controls, their mean ages being 58.8 years and 41.6 years, respectively. Additional data collected later showed nonfatally injured pedestrians to be intermediate in age, with a mean of 48.4 years. Thus, advancing age appeared to increase the pedestrian's risk both of being struck by a car and of dying once struck.

Regarding the effects of alcohol, significantly higher blood-alcohol concentrations were found in cases than controls. Appreciable increases in risk were noted even at the relatively low levels of 10 to 40 mg/100 cc. Putting together the age and alcohol data it appeared that there were two relatively discrete high-risk groups—the elderly who had been drinking little if any alcohol and the middle-aged who had been drinking heavily.

It was also found that the case group was more often foreign-born and of lower socioeconomic status than the controls, and less often married. However these differences could be explained by age differences between the case and control groups. Weather conditions, rain in particular, did not appear to be associated to any substantial degree with traffic deaths.

In addition to the case-control comparisons, information about the fatally injured group itself was of interest and importance. Only a small percentage lived outside of Manhattan and were commuters or out-of-town visitors. While the accidents were scattered about the city, most occurred outside of major business and shopping areas. The accidents occurred most frequently in the evening and night hours, suggesting the importance of emergency medical care during this time of day.

Evaluation and Role of the Case-Control Method

Case-control studies are the most readily and cheaply carried out of all analytic epidemiologic studies. For rare diseases they may be the only practical approach. Yet the problems involved in selecting appropriate control groups and collecting comparable information on cases and controls are often of such magnitude that the results of

case-control studies are open to a variety of legitimate questions and objections, generally more so than the results of prevalence and incidence studies.

Case-control studies have played a vital role in the development of many fruitful lines of study. For example the relationship of cigarette smoking to lung cancer was demonstrated in case-control studies before any incidence studies of this question were carried out. Because of their low cost, case-control studies should often be the first approach to the testing of a hypothesis. Similarly, they are useful for an exploratory study of a variety of variables (sometimes referred to as a "fishing expedition") to find clues and leads for further study.

REFERENCES

Haddon, W., Jr., P. Valien, J. R. McCarroll, and C. J. Umberger. 1961. A controlled investigation of the characteristics of adult pedestrians fatally injured by motor vehicles in Manhattan. *J. Chron. Dis.*, **14**:655–678.

MacMahon, B., and T. F. Pugh, *Epidemiology: Principles and Methods.* (Boston: Little, Brown, 1970), Chap. 12.

Sartwell, P. E., A. T. Masi, F. G. Arthes, G. R. Greene, and H. E. Smith. 1969. Thromboembolism and oral contraceptives: An epidemiologic case-control study. *Am. J. Epidemiology*, **90**:365–380.