

Chapter 10 - Learning Objectives

After completing this chapter, you should be able to:

1. Define the terms:

case-control study
case
control

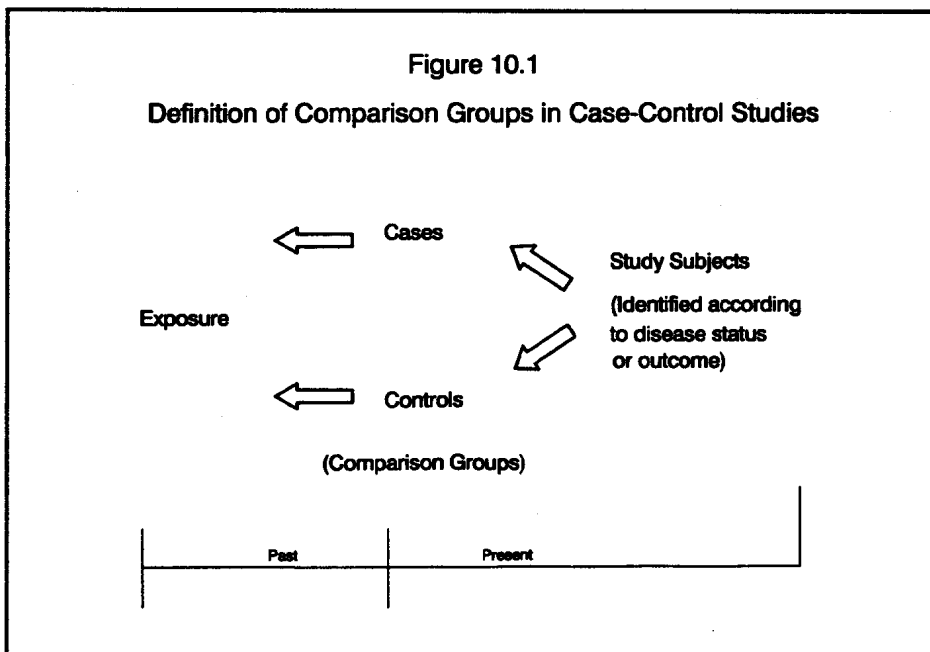
2. Describe how to select cases and controls.
3. Define the exposure variable in measurable terms.
4. Construct different measures of exposure.
5. Specify the data collection methods to be used in the study.
6. Describe types of bias.
7. Identify data analysis methods for case-control studies.
8. Describe why the measures of association used for cohort studies do not apply to case-control studies.
9. Construct analysis tables for a case-control study.
10. Calculate and interpret the odds ratio and the confidence interval for the odds ratio.
11. Interpret analysis tables for a case-control study.
12. Recognize advantages and disadvantages of case-control studies.
13. Design a hypothetical case-control study.

10 Case-Control Studies

Introduction

The case-control study is an analytic epidemiologic research design in which the study population consists of groups who either have or do not have a particular health problem or outcome. Comparison groups for case-control studies are formed on the basis of disease or health problem status. The study subjects with the health problem are called cases, and the study subjects without the health problem are called controls. In a case-control study we look back in time to measure exposure(s) of the study subjects (Figure 10.1). We compare the exposure among cases to the exposure among the controls to determine if the exposure could account for the health condition of the cases. The research hypotheses should clearly specify the expected relationship between the health problem and the exposure of interest.

Cases and controls defined



*Comparison
with cohort
study*

The case-control study differs from the cohort study in that individuals are categorized by their health problem or outcome; epidemiologists then look back in time to determine each individual's exposure status. In a cohort study, individuals are categorized by their exposure status; then epidemiologists look forward in time to determine the outcome. In Example 10.2 investigators examined the risk of myocardial infarction among women who were current users of oral contraceptives (OCs) and women who had discontinued using OCs (Slone et al., 1981). The cases were diagnosed with myocardial infarction during 1976 through 1979. The controls were admitted to the same hospitals during the same time period as the cases. The investigators looked back in time to measure exposure to contraceptives. Interviewers collected information about each woman's lifetime use of contraceptives, specifically OCs, before her diagnosis in 1976 through 1979.

Example 10.2

Oral Contraceptive Use and Myocardial Infarction

Problem: Is oral contraceptive (OC) use associated with myocardial infarction?

Research Hypothesis: Current OC use is positively associated with myocardial infarction.

Study Design: Multicenter hospital-based case-control study

Cases: Women who were hospitalized with a diagnosis of myocardial infarction (World Health Organization criteria).

Women 25 to 49 years old who were admitted to 155 hospitals in the northeastern United States between July 1, 1976, and June 30, 1979. Cases who had had a myocardial infarction, coronary-bypass surgery, cardiac-valve prosthesis, rheumatic heart disease, an occurrence of infarction after admission, or sickle-cell crisis preceding infarction were excluded from the study.

Controls: Women with no previous history of myocardial infarction who were admitted to the same hospitals and diagnosed during the same time period as the cases.

Four women, aged 25 to 49 years with diagnoses judged to be unrelated to contraceptive choice were selected at random for each case. The diagnoses

Example 10.2 (cont.)

considered unrelated to contraceptive choice included fractures, dislocations, soft-tissue injuries, gastroenteritis, appendicitis, disk disease, lower-back pain, renal calculus, hiatal hernia, reflux esophagitis, inguinal or femoral hernia, duodenal ulcer, gastric ulcer, and nontoxic thyroid nodules.

Exposure: Current OC use (OC use in the month before hospital admission).

Data Collection Methods: Hospital records were abstracted and cases and controls were interviewed by nurses in the hospital or at home after discharge.

(Slone et al., 1981)

Design and Data Collection Methods

Case Selection

The cases are the study subjects who have the health problem or outcome to be investigated. The definition of a case requires two specifications:

- An unambiguous and objective description of the health problem, including how the health problem is to be diagnosed (i.e., diagnostic procedures, laboratory tests, and clinical signs and symptoms),
- The eligibility criteria that will be used to select cases for the study (i.e., ages, no history of selected diseases and conditions, etc., and how, when, and where cases are identified) (Schlesselman, 1982). Cases may be all persons who are newly diagnosed with disease during a specified period of time (also known as incident cases).

Example 10.3 presents the case definition from a case-control study of pelvic inflammatory disease (PID) and current use of an intrauterine device (IUD) (Burkman et al., 1981). Clinically, cases were defined as women who received a hospital discharge diagnosis of PID. For analysis, the case definition was refined to

Case defined

Incident cases

Example 10.3

Use of Intrauterine Devices and Pelvic Inflammatory Disease

Problem: Is current intrauterine device (IUD) use associated with pelvic inflammatory disease (PID)?

Research Hypothesis: Current IUD use is positively associated with the development of PID.

Study Design: Multicenter hospital-based case-control study

Cases: Women with a hospital discharge diagnosis of PID. The diagnosis of PID was classified as *most certain*, *moderately certain*, and *least certain*. Women with anatomic findings of PID at laparoscopy, laparotomy, culdoscopy, or colpotomy (either the adnexa were visualized or an abscess cavity was drained) were classified as most certain. Women who were hospitalized with a diagnosis of PID for seven days or longer or who had a temperature greater than 100.4 degrees F on two or more days, and who did not undergo the operative procedures listed for the certain cases were classified as moderately certain. Women with a discharge diagnosis of PID who did not fit the criteria for the most certain and moderately certain cases were categorized as least certain.

Women 18 to 44 years old with a discharge diagnosis of PID between October 1976 and August 1978 were identified through a systematic review of admission lists, operative logs, ward census records, and discharge logs at one of 16 hospitals in nine cities in the United States. Women were excluded if they had not been sexually active during the previous year, had had sterilization surgery (or if partner had had sterilization surgery), had been pregnant during the previous six weeks, or had not menstruated in the past year.

Controls: Women who were admitted for acute conditions or elective surgery to the surgical and medical services of the same hospitals during the same time period as the cases.

Controls were selected using the same eligibility criteria as the cases, but they were not diagnosed with PID. Women with diagnoses of gallbladder disease; vascular, pelvic, or breast surgery; and other conditions that could influence contraceptive choice were excluded. Women with diagnoses of epilepsy, thromboembolism, cerebrovascular disease, or cancer were also excluded.

Exposure: Current IUD use (contraceptive method used during the 30 days before hospital admission). Women who had an IUD when admitted to the hospital or who had one removed within 30 days before being admitted to the hospital were classified as current IUD users.

Example 10.3 (Cont.)

Data Collection Methods: Information about the women enrolled in the study was abstracted from medical records, and a standardized interview was conducted during each woman's hospital stay. If a woman was discharged from the hospital before an interview was conducted, she was interviewed by telephone or in her home.

(Burkman et al., 1981)

assess the *certainty of diagnosis*. The investigators based their judgment on whether specific procedures necessary for confirmation of a diagnosis of PID had been performed and on the presence of selected signs and symptoms likely to indicate a diagnosis of PID. The criteria for defining cases of PID in this study were categorized as certain, moderately certain, and least certain.

Eligibility criteria for the cases in this study included a specific age range, sexual activity during the previous year, absence of sterilization surgeries (woman or her partner), menses during the past year, and no pregnancy during the six weeks before admission to the hospital. Study personnel described how, when, and where the cases were identified; admission logs were used to enroll women as potential cases. All women were interviewed before they were discharged from the hospital. Final eligibility was determined after discharge.

The eligibility criteria identify groups of women who could have experienced the exposure under study (Schlesselman, 1982). Women with conditions that might have limited their choices of contraception were excluded. For example, women who are not sexually active or who have had a tubal ligation are not likely to have recently used any contraceptive method including IUDs.

Good sources of cases include admissions and discharge logs in physician offices, clinics, and hospitals; surgical logs; disease or tumor registries; and for deaths, vital statistics records (Schlesselman, 1982). Potential sources of cases should be evaluated to understand the case identification process, the potential for identifying all cases, and the likely referral patterns for cases. For example, seriously ill cases who live a long distance

Sources of cases

from the medical care facility may be unable to travel to the facility, and thus may be missed as potential cases for study. Alternatively, cases from outside the defined study settings may be referred to the medical care facility for treatment and included for study. Failure to consider and manage these possibilities may bias risk estimates.

Control Selection

The controls are the study subjects who do not have the health problem or outcome under investigation. Controls should be similar to the cases with respect to the potential for exposure, since case-control analyses compare the rate of exposure among the cases to the rate of exposure among the controls (Schlesselman, 1982). The criteria used to select controls should be comparable in all ways with the criteria used to select cases, except the controls should not have the health problem under study. If controls are selected in this manner, differences in the rates of exposure are likely to reflect a true association between exposure and disease. If controls are not selected in this manner, then measures of the association between the health problem and the exposure may reflect the differences in the ways the cases and controls were selected and therefore may be biased.

The eligibility criteria for both cases and controls in Example 10.3 were similar and the likelihood that both groups of women were potential users of contraception was increased. Importantly, controls chosen from hospital admission records did not have diagnoses that were related to contraceptive use.

Alternatively, consider a case-control study in which the cases are identified from the discharge logs of a public hospital located in a large urban area. Assume that the controls are selected at random from the general population in the same urban area. These controls would be inappropriate because the general population probably includes many persons who would always use a private instead of a public hospital. These controls are not likely to be comparable to the cases with respect to access to medical care, diseases and conditions, socioeconomic status, education, and other factors.

The most common sources of control subjects are hospitals, clinics, physicians' offices, the community, or the general population. Selecting controls from the same hospital where cases are selected is often practical and cost efficient; these patients are referred to as hospital controls. However, hospital controls should be selected from

*Eligibility
criteria*

*Sources of
controls*

diagnostic categories that are unrelated to the exposure of interest. Controls may also be selected from the community or geographic area where the hospitalized cases reside; these controls are referred to as community or population controls. In some studies, neighbors, associates, friends, or relatives of cases have served as controls. Using the relatives of cases as controls will likely result in matching on variables such as diet, life-style behaviors, and family history of diseases.

Controls may be selected in several ways. Selection may involve sampling or may include the total population (excluding cases). Controls may be randomly or systematically selected from the total population. Finally, controls may be matched to cases on specific variables to make controlling for confounding in the analysis more efficient. Matching may be desirable if there is an unusual distribution of cases compared to controls with respect to a specific variable (e.g., age). Matching may be individual matching (also known as one-to-one matching) or frequency matching. If individual matching is used, one or more study subjects without disease is matched with a specific study subject who has disease. Alternatively, frequency matching involves matching several study subjects without disease in a given subgroup with study subjects who have disease.

*Sampling and
matching
controls*

Exposure Definition

The intensity of the exposure is assessed through measures related to frequency and time. Measures of exposure that relate to frequency may be dichotomous, polytomous, or continuous:

*Frequency
measures*

- Dichotomous: categorized as exposed or unexposed (e.g., never and ever diaphragm use)
- Polytomous: measured on more than two levels (e.g., never, occasional, and frequent condom use; number of 8-ounce drinks of alcohol per week; days per week when calcium is taken)
- Continuous: measured on the continuum of a unit of measure (e.g., age in years, parity, birthweight).

Time measures

Exposure may be intermittent (i.e., exposure to a risk factor is interspersed with periods of nonexposure) or continuous (i.e., exposure to a risk factor is constant where a single episode of exposure is not interrupted by exposure to an alternative risk factor or to no risk factor). Measures that relate to time include duration of exposure (e.g., total months of IUD use), time since first exposure (e.g., number of months since first OC use), time since last exposure (e.g., number of months since last pregnancy), and ages at first or last exposure. For some health problems, exposure before, during, or after selected reproductive events or events that are related to the health problem under study is important. For example, McPherson and colleagues (1983) suggested that OC use before the first term pregnancy might increase the risk of breast cancer. Exposure to alcohol, smoking, and various drugs during pregnancy has been related to adverse fetal, neonatal, and infant outcomes. Whether the exposure is current or past may also be important for studying selected health problems or outcomes. In Example 10.3, the definition of current exposure is precisely specified with respect to number of days and includes criteria about recent discontinuations.

Minimum exposure

Since measures of risk are expressed in terms of the exposure, minimum exposure must be defined. The cases and controls with the minimum exposure are the comparison or referent group. Minimum exposure may be defined as no exposure or as some low level of exposure judged to be insufficient to be related to disease.

Sources of exposure data

The primary sources of exposure data are 1) hospital or clinic records, 2) vital statistics, 3) employment, insurance, or social service records, and 4) direct contact with study subjects. Exposure data should be collected from existing documents when possible. When collecting exposure data, the same data collection methods should be used for both cases and controls. If existing sources are unavailable or inadequate, exposure data must be collected by contacting the study subjects directly (in person, by telephone, or by mail).

Data Collection Methods

In-person and telephone interview questionnaires, self-administered questionnaires, and forms designed to abstract information from medical records are used to collect and organize data for study and analysis. Careful form design and wording of the questions affect data quality. Questions used in other studies of the same health

problem or exposure are useful for planning and constructing questions for new studies. Using questions that provided valid data in other studies minimizes the time required to test question wording and promotes comparability of data across studies.

Questionnaire construction should begin with a list of variables needed for study. The list is best developed from a thorough review of the medical literature and should contain all the risk factors for the health problem and the exposure under study. A description of all possible ways to measure each variable should also be included. All data collection forms should be pretested, and data collection personnel (e.g., interviewers, abstractors) should be trained to use the forms correctly.

In Example 10.3, data collection involved medical record abstracts and personal interviews. The interview questionnaire contained questions to confirm the eligibility criteria, demographic questions, questions about factors that might be related to the occurrence of PID or to contraceptive choice, and histories of gynecologic conditions, chronic diseases, contraceptive use, pregnancies and menstrual events. (For additional sources on questionnaire design and data collection methods, see Bennett and Ritchie, 1975, Schlesselman, 1982, and Sudman and Bradburn, 1983.)

Bias

In the design of a case-control study, the methods by which study participants are selected and classified as diseased, not diseased, exposed, and unexposed affect validity (Schlesselman, 1982). Errors in design may cause the overrepresentation or underrepresentation of study participants into these categories and distort the measure of the association between the health problem and the exposure. Bias refers to the systematic errors that produce an inaccurate estimate of the association between the health problem and the exposure. Three types of bias can influence the results of the case-control investigation: selection bias, information bias, and confounding bias.

Selection bias. This type of bias refers to the process in which cases or controls are selected in a way that is related to exposure. Biased estimates of the association between the exposure and the

*Questionnaire
construction*

*Types of bias
defined*

Non-response bias

health problem can occur when the exposure is related to nonresponse, length of stay, survival, differential surveillance, diagnosis, referral, or selection of study participants.

Nonresponse bias is a type of selection bias that refers to the respondent's refusal or inability to participate in the study or to the field personnel's inability to contact potential study participants. Nonresponse bias occurs when the rate of exposure among nonparticipating cases is different from the rate of exposure among nonparticipating controls. While the study is being conducted, every effort should be made to locate study participants, to enlist their participation, and to minimize refusals. Interviewers may need training on how to locate difficult-to-find study participants. When refusal rates are high, several actions are needed to identify ways to reduce nonparticipation. Introductions and information about the study that is provided to study participants may need to be revised. Interviewers with high refusal rates may need specialized training about how to persuade reluctant participants to cooperate. Convening regular meetings for interviewers permits exchange of ideas about locating difficult-to-find participants and eliciting cooperation.

Refusals

However, if nonresponse is high at the conclusion of the study, some measure of the possible effects of nonparticipation on the results is needed. The investigator may perform a worst-case analysis in which all nonparticipating cases are assumed to be exposed and all nonparticipating controls are assumed to be unexposed. If the conclusions about the association between the health problem under study are unchanged with the worst-case analysis, nonparticipants are unlikely to have affected the results. The investigator may also compare characteristics of the participating and nonparticipating cases and controls.

Length of hospital stay bias

Bias related to length of hospital stay, another type of selection bias, may occur if cases are selected from a registry of current hospital patients instead of from admission or discharge logs. If cases are selected from a registry of current hospital patients, then cases who have been hospitalized for the longest period of time have a higher probability of being selected than cases admitted for minor conditions or cases who died. Furthermore, these cases may have other diseases and conditions that may be related to the disease or exposure under study. Therefore, use of admission or discharge logs for identifying potential study participants for hospital-based studies is preferable.

Two additional types of selection bias are survival bias and surveillance bias. Bias related to survival may occur if only the survivors of the outcome are selected as cases and if survival is related to the exposure of interest. Surveillance bias may occur when health problems that are mild or asymptomatic are diagnosed as a result of more frequent or thorough follow-up examinations and when more frequent or thorough examinations are conducted among study participants who have been subjected to suspected exposures. To assess whether surveillance bias is present in the data, cases and controls with frequent medical surveillance can be analyzed separately from cases and controls with infrequent surveillance.

Survival and surveillance bias

Diagnostic bias and referral bias are also types of selection bias. Diagnostic bias may occur if knowledge of the exposure status inappropriately alters the diagnosis. Assuming that diagnostic bias decreases as the severity of disease or the certainty of diagnosis increases, analysis of cases according to the certainty of diagnosis permits some assessment of the presence of this type of bias. Referral bias may occur if knowledge of the exposure status or if a variable (e.g., socioeconomic status) related to exposure inappropriately affects referral patterns. Referral bias is most likely to occur when a study is based in the hospital, clinic, or physician's office.

Diagnostic and referral bias

Potential biases introduced in selecting cases and controls may be reduced by designing the study so that selections are not necessary. The ideal study design involves the enrollment of all cases of a disease occurring in a defined geographic region during a specified period of time and the random selection of controls from the general population of the same area. Exclusions that are applied to the cases must also be applied to the controls. Since the purpose of the control group is to determine the rate of exposure expected in the case group, if no association between the exposure and the health problem is present, the controls should be comparable with the cases in all relevant ways except that they do not have the health problem.

Example 10.4 describes a population-based case-control study. The cases were all women diagnosed with histologically confirmed primary breast cancer who resided in one of eight geographic regions in the United States (The Cancer and Steroid Hormone Study, 1986). These women were identified through population-

Example 10.4

Oral Contraceptive Use and Breast Cancer

Problem: Is OC use associated with the risk of breast cancer?

Research Hypothesis: There is an association between OC use and breast cancer.

Study Design: Multicenter population-based case-control study

Cases: All women 20 to 54 years old with newly diagnosed breast cancer who reside in one of eight geographic regions in the United States. The women were identified through population-based tumor registries.

Controls: Women of the same ages selected at random from the general population of the same eight geographic regions during the same time period when cases were diagnosed.

Exposure: OC use before cancer diagnosis for cases and before interview for controls.

(The Cancer and Steroid Hormone Study, 1986)

based tumor registries, which are agencies that collect information on all new cases of cancer diagnosed in the region. The controls were selected at random from the general population of the same geographic regions as the cases. Since the controls were selected from the general population, the data on their contraceptive history will likely provide an accurate measure of the expected rate of exposure to oral contraceptives in these geographic regions. Although population-based case-control studies are ideal, they may not always be feasible. Identifying all case patients may be difficult if they do not routinely seek medical attention for the disease under study. Identifying controls from the general population may be expensive or logistically impossible. Furthermore, not all hospitals in the specified geographic regions or metropolitan areas will necessarily consent to participate in the study.

When studying certain mild conditions or outcomes for which only selected patients, such as more affluent patients, would be hospitalized, use of neighborhood or general population controls may pose problems that relate to selection bias. (Under these

circumstances, general population controls may not be comparable to cases, since controls with middle to lower socioeconomic status may never be hospitalized for the mild condition.) Preferably, the controls should be restricted to individuals who would likely have been hospitalized if they developed the condition under study.

Hospital-based case-control studies maximize similarities between cases and controls. In the hospital-based case-control study, controls should be selected from patients admitted to the same hospitals for other minor conditions that are known not to be associated with the exposure. If such measures are taken, cases and controls are more likely to have similar socioeconomic status, life-style behaviors, religion, and other traits that can determine patterns of hospitalization. Although this selection procedure promotes comparability between cases and controls, the results may not be generalizable to the entire population.

If the control group is chosen from hospitalized patients, it seems best to include individuals with many different illnesses, none of which is known to be associated with the exposure being studied. In this way, even if one of the diseases is found to be associated with the exposure under study, it will likely have little effect on the study results since the number of patients with one specific illness is small (Example 10.5).

Example 10.5

Selection Bias in Hospital-Based Studies

One of the first case-control studies of OC use and cardiovascular disease included many women with gallbladder disease in the control group. Shortly after the study was concluded, several reports indicated that women using OCs were at increased risk of gallbladder disease. Thus the control group contained women who had developed gallbladder disease because they had used OCs. As a result, a spuriously high rate of exposure to oral contraceptives was noted among the controls, and the study artificially underestimated the relative risk of cardiovascular disease in oral contraceptive users.

(Boston Collaborative Drug Surveillance Program, 1973)

Information bias. This type of bias refers to the collection of incorrect information about exposure that results in an incorrect measurement of the exposure. Obtaining an accurate exposure history from the cases and controls is one of the major difficulties in conducting a valid case-control study. Participants' recall of past exposures may be inconsistent and inaccurate. Exposure information may be difficult to remember, particularly when the exposures occurred in the distant past. To improve recall, memory aids such as calendars, diaries, photographs, or other visual materials may be helpful. Exposure information should be validated whenever possible, using sources independent of the study subject's report (Example 10.6).

Example 10.6

The Use of Memory Aids to Reduce Information Bias

In studies of the association between oral contraceptive use and reproductive cancers, accurate information on use of specific formulations of oral contraceptives and use in the distant past was essential. Researchers in Great Britain designed a study to compare self-reported oral contraceptive histories collected during personal interview with data collected prospectively from the Oxford Family Planning Association cohort study. Two memory aids were used to assist recall during interview: a calendar of important life events to which the study participant might relate her use of oral contraceptives, and a book of color photographs of all oral contraceptive preparations marketed in Great Britain. Interview data collected with the calendar and photographs consistently had better agreement with the cohort records data than interview data collected without the aids. Agreement between self-reported data and cohort records within six months was 56.2% for total duration, 81.2% for date of first use, 66.7% for date of last use, and 79.2% for date of use prior to first term pregnancy. For interviews conducted without the memory aids, the rate for total duration was 39.2%, for date of first use, 80.4%, for date of last use, 52.9%, and for date of use prior to first term pregnancy, 70.6%.

(Coulter et al., 1986)

Recall bias

More important than forgetfulness, however, is whether cases remember exposures to risk factors in the past differently than controls. Recall bias refers to the effect of cases remembering exposures differently than controls. In some case-control studies,

recall bias may be of such concern that it influences the choice of a control group (Example 10.7).

Example 10.7

The Effect of Potential Recall Bias on the Choice of Controls

Investigators of congenital malformations seek information on exposures early in pregnancy. Because of the psychological trauma of giving birth to a child with defects, mothers of cases may be more likely than mothers of controls to recall exposures to potential risk factors during pregnancy; they may, in fact, search for an explanation of the malformation. This enhancement of recall could result in bias, and spuriously create a difference between the cases and the controls in their reported exposure histories. Recall bias in studies of malformations may be minimized by choosing control infants with different birth defects than the birth defect of interest.

In a case-control study of the relationship between birth defects and maternal consumption of caffeine, six groups of case patients (inguinal hernia, cleft lip, cardiac defect excluding murmur, pyloric stenosis, isolated cleft palate, and neural tube fusion defect) were compared to all other malformed infants. Mothers of malformed infants who did not have one of the six case defects would likely have similarly enhanced recall as mothers of the case patients.

(Rosenberg et al., 1982)

A similar rationale has been used in the defense of hospital-based case-control studies. Since the controls in hospital-based studies are also ill, they would be expected to have a similar enhancement of memory as the cases.

Confounding bias. This bias refers to the effect of an extraneous factor that distorts an apparent association between the health problem and the exposure under study or that obscures an underlying true association (Schlesselman, 1982). That is, the association between an exposure and the health problem under study may actually be due to another variable. Alternatively, the lack of an association could result from failure to control for the effect of some extraneous factor. To be a confounder, a variable must be associated with, but not a consequence of, the exposure, and be a risk factor for the health problem under study

(Schlesselman, 1982). Confounding may be controlled in the data analysis phase if the relevant information has been collected.

Data Analysis Methods

Analyses should be planned concurrently with designing the study protocol and data collection form or questionnaire to ensure that information is collected for all important variables in a manner that is appropriate for analysis. The first steps of the analysis begin with organizing the data into tables that are used to make comparisons between cases and controls. Typical tables presented for the analysis of case-control studies include characteristics (demographics and risk factors) of the cases and controls (Table 10.8), information about the magnitude of the association between the health problem and the exposure under study (Tables 10.14 and 10.16), and information about the risk of the health problem according to various subgroups of the cases and controls (Table 10.18).

Analysis Table of Characteristics of Cases and Controls

Careful examination of differences of various characteristics between cases and controls provides important information about the comparability between the case patients and the controls. First, these data permit comparisons with studies published in the literature. The risk factors present in the data under study can be compared with known risk factors in other studies. When known risk factors are not present in the data under study, the researcher should evaluate and question his data collection procedures and findings. Second, analysis of known risk factors may provide information about which variables will potentially confound an association between the health problem and the exposure under study. Third, evaluation of similarities or dissimilarities between the cases and controls may provide information that may indicate if cases and controls were selected or interviewed in a comparable manner. For example, in multicenter studies, are the geographic, hospital, or clinic distributions of the cases and controls similar?

Table 10.8 presents selected percentage distributions of demographic characteristics and risk factors for cases and controls enrolled in the case-control study of oral contraceptive use and

Table 10.8

**Characteristics Women with
Epithelial Ovarian Cancer and Controls**

<u>Characteristics</u>	<u>Cases (n = 492) %</u>	<u>Controls (n = 4,228) %</u>
Age (years)		
20-29	11.0	6.5
30-39	19.9	23.4
40-49	35.8	42.0
50-54	33.3	28.1
Parity		
Nulliparous	25.2	14.8
1	14.4	11.0
2	26.2	24.2
3	19.1	21.8
4+	14.6	27.7
Unknown	0.4	0.5
Obesity [(Weight (g)/Height ² (cm))]		
<2.25	48.8	56.0
2.25-2.49	26.2	21.5
2.50+	24.8	22.1
Unknown	0.2	0.4

(The Cancer and Steroid Hormone Study, 1987)

epithelial ovarian cancer (The Cancer and Steroid Hormone Study, 1987). Cases were more likely than controls to be younger, have lower parity, and be obese.

The Odds Ratio as an Estimate of the Relative Risk

Data collected from case-control studies do not permit a direct calculation of the relative risk. Because the proportion of the population that the cases and the controls represent is usually not known, we cannot derive incidence rates in the exposed and unexposed populations. Instead, the odds ratio is used to estimate

*Rare disease
assumption*

the relative risk. When the health problem is rare among the exposed and unexposed persons in the general population, the odds ratio closely approximates the relative risk. Historically, case-control studies have been the preferred study design when the incidence of the study outcome is rare, such as in cancer studies. Example 10.9 and Table 10.10 permit the comparison between the use of the relative risk in a hypothetical cohort study and the use of the odds ratio in a case-control study with similar hypotheses.

Example 10.9

Oral Contraceptive Use and Cervical Intraepithelial Neoplasia Cohort Study

Problem: Is OC use related to the risk of cervical intraepithelial neoplasia (CIN)?

Research Hypothesis: OCs are associated with the risk of CIN.

Study Design: Multicenter cohort study

Exposed: Women 18 to 44 years old who were using OCs when they visited any of 20 family planning clinics.

Not Exposed: Women of the same ages who were not using OCs when they visited the same family planning clinics.

Outcome: Exposed and unexposed women were followed for one year to observe occurrences of biopsy-proven CIN.

Data Collection Methods: Clinic records for the current visit were abstracted for each visit during one-year follow-up.

Clinic records indicated that the one-year incidence of cervical dysplasia among the exposed women would be 5% and unexposed women would be 2%. The cumulative incidence relative risk (Table 10.10) was:

$$\text{CIR} = \frac{a/n_1}{b/n_0} = \frac{100/2,000}{40/2,000} = 2.5$$

Table 10.10
Analysis Table

Cohort Study

	OC Use	
	Yes	No
<u>Cervical Dysplasia</u>		
Yes	a 100	b 40
No	c 1,900	d 1,960
Total	n_1 2,000	n_0 2,000

Case-Control Study

	OC Use		
	Yes	No	Total
<u>Cervical Dysplasia</u>			
Yes (Case)	100	40	140
No (Control)	69	71	140

If this problem had been investigated in a case-control study (Table 10.10), the 140 cases of cervical dysplasia would constitute the group of cases. An equal number of controls would have been selected at random from the $1,900 + 1,960 = 3,860$ women who did not have cervical dysplasia ($69 = 1,900 * 140/3,860$ controls who had used OCs, and $71 = 1,960 * 140/3,860$ controls who had never used OCs). Information about exposure to oral contraceptives would have been collected through in-person interviews with the study participants.

The odds ratio was:

$$OR = \frac{a/b}{c/d} = \frac{100/40}{69/71} = \frac{100 * 71}{69 * 40} = 2.6$$

Because the incidence of cervical dysplasia was low in the general population, the case-control study yielded an odds ratio of 2.6. This ratio is approximately equal to the cohort study cumulative incidence relative risk of 2.5. However, the case-control design required a substantially smaller study size (280) than the cohort study (4,000).

Alternatively, if the incidence of the health problem under study is high, the odds ratio will be an inadequate estimate of the relative risk (Example 10.11). In the cohort design, the proportion of the exposed with the outcome (1,000/2,000) is the same as the proportion of cases exposed in the case-control study (1,000/2,000). A similar relationship exists among the unexposed women in the cohort study (1,000/10,000) and the controls who were exposed in the case-control study (200/2,000). The unbiased estimate of the cumulative incidence relative risk from the cohort study is CIR=5.0. Because the incidence of the outcome was high in the general population the estimate of relative risk from the odds ratio, OR=9.0, is a poor estimate.

Case-Control Study Analysis Table for the Odds Ratio and Confidence Interval

Odds ratio

In case-control studies, the odds ratio (OR) estimates the relative risk as a measure of the magnitude of the association between the exposure and the health problem under study. Table 10.12 presents the analysis table for the odds ratio. Table 10.12 is similar to the tables used to estimate the relative risk (RR) and the incidence density ratio in cohort studies and randomized clinical trials. However, the risk measures applicable for cohort studies cannot be calculated from data gathered by the case-control design because, in a case-control study, the health problem or outcome is sampled, not the exposure. The ratio of risks in a case-control study may be estimated *by dividing the odds of disease among the exposed (a/c) by the odds of disease among the unexposed (b/d)* (See 10.12.1). An OR larger than 1.0 implies that the risk of the health problem is increased if the study participant was exposed. An OR smaller than 1.0 implies protection as a result of the exposure (i.e., the risk of the health problem is decreased if the study participant was exposed).

Example 10.11**High Incidence Rates
Analysis Table****Cohort Study**

<u>Outcome</u>	<u>Exposure</u>	
	<u>Yes</u>	<u>No</u>
Yes	1,000	1,000
No	1,000	9,000
Total	2,000	10,000

$$\text{CIR} = \frac{1,000/2,000}{1,000/10,000} = \frac{.5}{.1} = 5.0$$

Case-Control Study

<u>Outcome</u>	<u>Exposure</u>		
	<u>Yes</u>	<u>No</u>	<u>Total</u>
Yes (Case)	1,000	1,000	2,000
No (Control)	200	1,800	2,000
Total	1,200	2,800	

$$\text{OR} = \frac{1,000 * 1,800}{1,000 * 200} = 9.0$$

Table 10.12

Case-Control Study Analysis Table for Estimating the Odds Ratio

<u>Outcome</u>	<u>Exposure</u>		<u>Total</u>
	<u>Number Exposed</u>	<u>Number Not Exposed</u>	
Outcome Present (Cases)	a	b	m ₁
Outcome Absent (Controls)	c	d	m ₀
Total	n ₁	n ₀	t

Odds Ratio (OR) = $\frac{\text{The Odds of Disease Among the Exposed}}{\text{The Odds of Disease Among the Unexposed}}$

(10.12.1)

$$= \frac{a/c}{b/d} = \frac{a*d}{b*c}$$

Confidence interval for odds ratio

The confidence interval (CI) for the odds ratio (Mantel and Haenszel, 1959, Miettinen, 1976) is given by the formula 10.12.2.

(10.12.2)

Confidence Interval for OR = $OR^{(1 \pm z/\chi)}$

where z is a normal variate, and

(10.12.3)

$$\chi = \sqrt{\chi_{MH}^2}, \text{ and } \chi_{MH}^2 = \frac{(t - 1) * (a * d - b * c)^2}{n_1 * n_0 * m_1 * m_0}$$

This confidence interval provides a good estimate when the OR is close to 1.0 but becomes less stable for ORs greater than, say 3. Table 10.13 presents the risk of epithelial ovarian cancer among

women who have ever used oral contraceptives (The Cancer and Steroid Hormone Study, 1987). Women who have ever used OCs have a 40 percent $[(1 - 0.6) * 100]$ decreased risk of developing epithelial ovarian cancer than women who have never used OCs. Oral contraceptive use protects against ovarian cancer.

Using the data in Table 10.13, the confidence interval for the odds ratio is computed by performing the following steps:

Step 1:

$$\begin{aligned}\chi_{MH}^2 &= \frac{(t - 1) * (a * d - b * c)^2}{n_1 * n_0 * m_1 * m_0} \\ &= 4,719 * \frac{(250 * 1,532 - 242 * 2,696)^2}{2,946 * 1,774 * 492 * 4,228} \\ &= 31.51\end{aligned}$$

$$\begin{aligned}\chi_{MH} &= \sqrt{\chi_{MH}^2} \\ &= 5.61\end{aligned}$$

Step 2: For 95% confidence interval, $z = 1.96$

Step 3:

$$\begin{aligned}\text{Lower limit} &= OR^{(1 - z/\chi)} \\ &= e^{[\ln OR * (1 - z/\chi)]} \\ &= e^{[\ln 0.6 * (1 - 1.96/5.61)]} \\ &= 0.5\end{aligned}$$

$$\begin{aligned}\text{Upper limit} &= OR^{(1 + z/\chi)} \\ &= e^{[\ln OR * (1 + z/\chi)]} \\ &= e^{[\ln 0.6 * (1 + 1.96/5.61)]} \\ &= 0.7\end{aligned}$$

Because the confidence interval does not include 1.0 in Table 10.13, the odds ratio is considered statistically significant. That is, OC use protects against ovarian cancer, and the estimate of the relative risk is significantly less than 1.0.

Table 10.13

Oral Contraceptive Use and Epithelial Ovarian Cancer

<u>Epithelial Ovarian Cancer</u>	<u>Ever OC Use</u>		
	<u>Yes</u>	<u>No</u>	<u>Total</u>
Yes	250	242	492
No	2,696	1,532	4,228
Total	2,946	1,774	
OR = 0.6 (95% CI: 0.5 - 0.7)			

(The Cancer and Steroid Hormone Study, 1987)

In Example 10.14, compared with women who have no history of legal abortion, women with a previous legal abortion have a relative risk of 1.4 of having placenta previa in a later pregnancy (Grimes and Techman, 1984). In this analysis, because the confidence interval includes 1.0, the investigators concluded that legal abortion appeared to have little effect on the development of placenta previa in a later pregnancy.

Evaluating some frequency or time-related measure of the exposure may be important for some health problems. That is, does the risk of the health problem differ for different levels of exposure? Does the risk of the health problem increase (decrease) with increasing (decreasing) levels of the exposure? Table 10.15 presents the risk of ovarian cancer according to years duration of OC use (The Cancer and Steroid Hormone Study, 1987). An odds ratio is computed for each duration grouping of years of OC use, with never users as the referent group. The risk of ovarian cancer decreased from 0.6 among women who use OCs for less than one year to 0.3 among women who use OCs for at least 10 years.

Other possible measures for evaluating the association between ovarian cancer and OC use might include years since first (last) use of OCs, micrograms of estrogen (milligrams of progestin) in the oral contraceptive used, etc.

Example 10.14**Legal Abortion and the Risk of Placenta Previa**

Problem: Is legal abortion associated with placenta previa in a later pregnancy?

Research Hypothesis: Legal abortion increases the risk of placenta previa in a later pregnancy?

Study Design: Hospital-based case-control study

Cases: Women who gave birth to infants weighing >499 g at Grady Hospital (Atlanta, Georgia) between January 1, 1975 and December 31, 1979. The women were diagnosed either with complete placenta previa (documented by ultrasonography) that necessitated cesarean delivery of the infant or with placenta previa (documented by ultrasonography) that led to bleeding and required hospitalization during pregnancy.

Controls: Women without placenta previa who delivered infants weighing >499 g at Grady Hospital during the study period; they were randomly selected.

Exposure: A history of one or more previously induced abortions.

Data Collection Methods: At the first prenatal visit, information about the cases and controls was collected through personal interview; and abstracts of computerized medical records and charts were used to collect information for the remainder of the pregnancy. Data obtained at each interview was validated, when possible, through linkage of family planning clinic records.

Results:

<u>Outcome</u>	<u>Exposure</u>	
	<u>Previous Legal Abortion</u>	<u>No Previous Legal Abortion</u>
Placenta Previa		
Yes	12	56
No	9	59

OR = 1.4 (95% CI: 0.5 - 3.6)

(Grimes and Techman, 1984)

Table 10.15**The Risk of Epithelial Ovarian Cancer by
Duration of Oral Contraceptive Use**

<u>Duration of Use (Years)</u>	<u>Cases</u>	<u>Controls</u>	<u>Crude Odds Ratio</u>	<u>95% Confidence Interval</u>
Never	242	1,532	1.0	Referent
< 1	40	414	0.6	(0.4 - 0.9)
1-2	65	602	0.7	(0.5 - 0.9)
3-4	40	397	0.6	(0.4 - 0.9)
5-9	39	594	0.4	(0.3 - 0.6)
≥ 10	13	328	0.3	(0.1 - 0.4)

(The Cancer and Steroid Hormone Study, 1987)

Controlling for Confounding in the Analysis

Unlike selection and information bias, we can adjust for confounding in the analysis phase. During data collection, we should collect information on potentially confounding variables so that they can be addressed during analysis. To control for confounding, we arrange the data in strata according to levels of the potentially confounding variable. Then, the Mantel-Haenszel method may be used to compute a summary (adjusted) relative risk and summary χ^2 (Mantel and Haenszel, 1959). The adjusted relative risk is a weighted relative risk that takes into account the relative risk estimates for each stratum.

To determine whether or not a variable confounds the association between the outcome and the exposure, the crude (unadjusted) relative risk (Formula 10.12.1) is compared with the adjusted relative risk. If the adjusted relative risk estimate differs from the crude relative risk by more than some previously specified percentage, the variable may be considered a confounder. Under these circumstances, the adjusted relative risk estimate should be presented instead of the crude estimate.

In Table 10.16, the relative risk adjusted for age (3.1) differs from the crude relative risk (2.4) by 29%. This result suggests that

Table 10.16**Current Oral Contraceptive Use and Myocardial Infarction
Adjustment for Confounding by Age**

<u>Age</u>		<u>OC Use</u>		<u>OR</u>	<u>95% CI</u>
		<u>Yes</u>	<u>No</u>		
<40	Cases	21	26		
	Controls	17	59	2.8	(1.3 - 6.1)
40-44	Cases	8	44		
	Controls	2	50	4.5	(1.0 - 20.3)
Total	Cases	29	70		
	Controls	19	109		

Crude RR = 2.4 (95% CI: 1.2 - 4.5)

RR_{MH} = 3.1 (95% CI: 1.6 - 6.3)

(Hogue et al., 1985)

age is a confounding variable in this study. Because age confounds the association between current OC use and the risk of myocardial infarction, the adjusted relative risk (3.1) should be reported. Current OC use is associated with a statistically significant increased risk of myocardial infarction; overall, women who currently use OCs have three times the risk of myocardial infarction as women who are not currently using OCs.

Effect Modification Analysis Table

Some health problems or outcomes require analyses of subgroups of study participants. That is, does the risk of the health problem for the exposure of interest vary for different ages or is the risk greater for some racial or ethnic groups? Do women with a particular characteristic have a different risk than women without the characteristic? When the association between a health problem and an exposure varies according to categorizations of a particular variable, that variable is considered an effect modifier. In Table 10.16, age is not only a confounder but also an effect modifier. The relative risks differ according to age. Among women who are 40 to 44 years old, current OC users have 4.5 times the risk of a myocardial infarction as women of the same ages who were not currently using OCs. Among younger women the risk is 2.8.

Table 10.17 presents the risk of developing epithelial ovarian cancer among women who have used OCs and women who have never used OCs, by age and parity (Cancer and Steroid Hormone Study, 1987). The risk of epithelial ovarian cancer does not appear to vary by age among women who had ever used OCs and women who had never used OCs. However, the risk of epithelial ovarian cancer varied slightly for different levels of parity. For example, nulliparous women had a risk of 0.7, women with parity 1 to 4 had a statistically significant reduced relative risks of 0.4 or 0.5, and women with parity 5 or more had a relative risk of 1.2.

Although effect modification seems similar to confounding, the two are very different. A variable can be an effect modifier of the association between the outcome and an exposure, a confounder of the association, both, or neither. When effect modification appears to be present, Woolf's χ^2 test for heterogeneity may be used to test whether the variations in stratum-specific relative risks are real or due to chance (Woolf, 1955).

Advantages and Disadvantages

Advantages:

- Case-control studies are useful for studying health problems that occur infrequently.

Table 10.17

**Oral Contraceptive Use and the Risk of Epithelial
Ovarian Cancer by Age and Parity**

<u>Characteristic</u>	<u>Cases</u>	<u>Controls</u>	<u>Odds Ratio*</u>	<u>95% Confidence Interval</u>
Age (years)				
<30	15/ 32 [†]	53/206 [†]	0.5	(0.3 - 1.1)
30 - 34	114/ 40	695/414	0.6	(0.4 - 0.9)
35 - 39	61/ 29	454/515	0.4	(0.3 - 0.7)
40 - 44	33/ 33	199/474	0.4	(0.3 - 0.7)
45 - 49	10/ 33	81/462	0.6	(0.3 - 1.2)
50 - 54	9/ 38	50/351	0.6	(0.3 - 1.3)
Parity				
Nulliparous	63/ 54	231/302	0.7	(0.4 - 1.0)
1	30/ 34	135/292	0.5	(0.3 - 0.9)
2	60/ 53	324/672	0.4	(0.3 - 0.6)
3	52/ 37	305/578	0.4	(0.2 - 0.6)
4	22/ 10	245/294	0.4	(0.2 - 0.8)
5+	15/ 17	292/284	1.2	(0.6 - 2.4)

* In each stratum, ever users are compared to never users.

† Did not use/used oral contraceptives.

(The Cancer and Steroid Hormone Study, 1987)

- Case-control studies are useful for studying health problems with a long latent interval.
- The relatively short study period required usually makes case-control studies less time consuming and less expensive than cohort studies.
- Case-control studies are useful for characterizing the effects of a variety of potential risk factors on the health problem under study.

Disadvantages:

- Because cases and controls may be selected from two separate populations, it is difficult to ensure they are comparable with respect to extraneous risk factors and other sources of bias.
- Exposure data are collected from records or by recall after the disease has occurred. Records may be incomplete, and recall of past events is subject to human error and the possibility of selective recall.
- Case-control studies cannot be used to determine incidence rates.
- If the health problem is relatively common in the population (i.e., > 5%-10%), the odds ratio is not a reliable estimate of the relative risk.
- Case-control studies cannot be used to determine the other possible health effects of an exposure. By definition, case-control studies are concerned with only one outcome.

Practice Exercises

1. In Example 10.3, investigators examined the association between IUDs and PID. Using the information presented in Table 10.18, analyze and interpret the data.

Table 10.18

**The Risk of Pelvic Inflammatory Disease Among Women
Who Use Intrauterine Devices and
Women Who Use No Method of Contraception**

<u>PID</u>	<u>IUD Use</u>	
	<u>Current IUD User</u>	<u>No Method Currently</u>
Yes	841	724
No	518	967

(Burkman et al., 1981)

- (a) Compute the odds ratio and the 95% confidence interval for this study.

- (b) Interpret the odds ratio and the confidence interval.

Case-Control Studies

(c) What are some questions the investigators should consider regarding the choice of the control group?

(d) What are some other potential risk factors for PID that need to be considered in the analysis?

2. Example 10.19 presents a case-control study designed to investigate whether early age at first coitus is a risk factor for cervical cancer.

Example 10.19

Early Age at First Coitus and Cervical Cancer

Problem: Is age at first coitus associated with cervical cancer?

Research Hypothesis: Age ≤ 15 years at first coitus is positively associated with subsequent development of cervical cancer.

Study Design: Hospital-based case-control study

Cases: Women of any age with a histologic diagnosis of invasive cancer of cervix at one hospital.

Controls: Women of any age who were healthy and attended the hospital's family planning clinic.

Exposure: The woman's report that her age at first coitus was ≤ 15 years.

Data Collection Methods: Questionnaires were administered to cases and controls.

Results:

Cervical Cancer	<u>Age at First Coitus</u>	
	<u>≤ 15 years</u>	<u>>15 years</u>
Yes (Cases)	36	78
No (Controls)	11	95
Crude OR = 4.0 (95% CI: 2.0 - 8.1)		

(Andolusi, 1977)

- (a) Interpret the odds ratio and confidence interval.

Case-Control Studies

3. Example 10.20 presents a study of OC use and the risk of PID.

Example 10.20

Current Oral Contraceptive Use and Pelvic Inflammatory Disease

Problem: Is current use of OCs associated with pelvic inflammatory disease (PID)?

Research Hypothesis: Current OC use is associated with PID.

Study Design: Hospital-based case-control study

Cases: Women 18 to 44 years old who were admitted to nine hospitals in the United States with an initial episode of PID.

Controls: Women 18 to 44 years old with no history of PID who were admitted to the same hospitals as the case patients but had acute conditions or elective procedures not related to PID.

Women in either group who reported sterility, recent pregnancy, or lack of sexual activity, as well as women with conditions that might contraindicate OC use, were excluded from the study.

Exposure: The woman's report of the contraceptive method used in the three months prior to interview.

Data Collection Methods: Standard questionnaires were administered to both cases and controls.

Results:

<u>Outcome</u>	<u>Exposure</u>		<u>Total</u>
	<u>Current OC User</u>	<u>No Method</u>	
PID			
Yes	139	170	309
No	831	558	1,389
Total	970	728	

(Rubin et al., 1982)

(a) Compute the odds ratio and the 95 % confidence interval.

(b) Interpret the odds ratio and confidence interval.

Case-Control Studies

4. In Example 10.21, the description of a study of depo-medroxyprogesterone acetate (DMPA) and the risk of breast cancer is presented:

Example 10.21
DMPA Use and Breast Cancer

Problem: Is DMPA use associated with the risk of breast cancer?

Research Hypothesis: DMPA is associated with the risk of breast cancer.

Study Design: Case-control study

Cases: Women 25 to 58 years old who were diagnosed with breast cancer between January 1, 1982 through March 31, 1984. The women were retrospectively enrolled using the National Tumor Registry records in Costa Rica.

Controls: Women 25 to 58 years old who were selected at random from the general population with a multistage probability household survey.

Exposure: The woman's report of using an injectable contraceptive.

Data Collection Methods: Personal interviews conducted in the home and tumor registry records were abstracted.

Results:

<u>Breast Cancer</u>	<u>DMPA Use</u>	
	<u>Yes</u>	<u>No</u>
Cases	19	129
Controls	49	724

Crude OR = 2.2 (95% CI: 1.3 - 3.8)

(Lee et al., 1987)

- (a) Interpret the odds ratio and confidence interval.

5. Circle true (T) or false (F).
- (a) T/F In the case-control study, study subjects from the study population are assigned to treatment or comparison groups.
 - (b) T/F Cases are those who have the health problem under study.
 - (c) T/F The purpose of the case-control study is to decide whether the exposure under study could account for the health condition of the cases.
 - (d) T/F If there is an association between the exposure and the health problem, then the control group provides an estimate of the rate of exposure in the cases.
 - (e) T/F To be enrolled in a case-control study, a case must clearly have the health problem under study.
 - (f) T/F Controls may be matched to cases to adjust for potentially confounding variables.
 - (g) T/F The effects of variables on which subjects have been matched cannot be evaluated.
 - (h) T/F Cases and controls should be matched on variables related to exposure.
 - (i) T/F Variables used for matching during data collection need not be considered as matched variables for analysis.
 - (j) T/F Relatives of cases may not be used for matching.
 - (k) T/F The exposure variable must be defined in clear, measurable terms.

Case-Control Studies

6. Multiple Choice. Select one response.

6.1 All of the following are frequency measures of exposure except?

- (a) Days per week vitamins taken
- (b) Total months of condom use
- (c) Parity
- (d) Ever use of natural methods of birth control
- (e) History of previous abortion

6.2 All of the following are measures of exposure related to time except?

- (a) Age at first intercourse
- (b) Time since IUD was inserted
- (c) Current IUD use
- (d) Duration of OC use
- (e) Ever breast-fed

7. Which of the following are appropriate (True) methods to help reduce bias? Circle true (T) or false (F).

- (a) T/F Select all cases in a defined region.
- (b) T/F Choose controls at random from an appropriate population.
- (c) T/F If you use a hospital-based design, then the control group should include individuals with different illnesses not associated with the exposure.
- (d) T/F Use memory aids to improve subject recall.
- (e) T/F Select controls whose recall bias is likely to be similar to the recall bias of the cases.
- (f) T/F Validate recall with medical or other institutional records whenever possible.
- (g) T/F Omit from the study any participants with poor memory.

8. Multiple Choice. Select one response.

8.1 All the data sources listed below are useful in case-control studies except:

- (a) Medical events that occur after diagnosis of the health problem of interest
- (b) Abstracts of clinical records
- (c) Vital statistics
- (d) Physical examination
- (e) Questionnaire data

8.2 Why do the measures of association used for cohort studies not apply to case-control studies?

- (a) Because you select controls from a different population than cases
- (b) Because of selection bias
- (c) Because you cannot determine rates without sampling exposure
- (d) Because you sample exposed and unexposed individuals in a case-control design

8.3 All of the following are types of bias except:

- (a) Selection
- (b) Surveillance
- (c) Effect modification
- (d) Confounding
- (e) Diagnostic

Case-Control Studies

9. Circle true (T) or false (F).

- (a) T/F Case-control studies are appropriate for infrequently occurring health problems.
- (b) T/F Case-control studies are appropriate for determining incidence rates.
- (c) T/F Case-control studies are inappropriate for studying problems with long latency.
- (d) T/F Case-control studies are generally less expensive than cohort studies.
- (e) T/F In case-control studies it is often difficult to ensure that cases and controls are comparable.
- (f) T/F Data collected for case-control studies permit calculation of relative risk.
- (g) T/F The odds ratio is an appropriate estimate of the relative risk when the incidence rates of the health problem under study are high.
- (h) T/F Case-control studies usually require smaller study sizes than cohort studies.
- (i) T/F A variable may be an effect modifier or a confounder, but not both.
- (j) T/F Summary relative risk measures may mask increases or decreases in risk that are present in subgroups.
- (k) T/F An OR greater or smaller than 1.0 implies that the risk of the health problem is associated with the exposure.
- (l) T/F Confounding bias may be minimized or eliminated by exclusions, matching, or analytic methods.

10. In outline form, design a case-control study based on the following problem situation. State the problem and the hypothesis. Describe the cases and the controls and how both groups of women will be enrolled in the study. Define the exposure in measurable terms. Specify the data collection methods. Develop the analysis table for measuring any possible association between the outcome and the exposure.

Background: As a researcher you have a collaborative relationship with several health care providers that allows you to conduct scientific studies. In particular, you want to investigate the risk of myocardial infarction among women who used OCs in the past.

Problem:

**Research
Hypothesis:**

Study Design:

Cases:

Case-Control Studies

Controls:

Exposure:

Data Collection Methods:

**Data
Analysis:**

11. Table 10.22 presents results from a study of ectopic pregnancy and current contraceptive methods (Ory et al., 1981).

<u>Contraception and Ectopic Pregnancy</u>				
<u>Current Method</u>	<u>Women With Ectopic Pregnancy</u>	<u>Non-Pregnant Controls</u>	<u>Odds Ratio</u>	<u>95% CI</u>
IUD	67	497		
Barrier or natural methods	57	573		
OC	32	775		
No method	319	1078		

- (a) Compute the odds ratios and confidence intervals for current IUD, barrier or natural methods, and oral contraceptive use. Consider women who used no method as the referent category.

Suggested Answers to Practice Exercises

1. Using the information in Table 10.18, analyze and interpret the data.

(a)

$$\text{OR} = \frac{841/724}{518/967} = 2.168$$

$$n = 3050$$

$$\chi^2 = \frac{3049 * (841 * 967 - 724 * 518)^2}{(841 + 518) * (724 + 967) * (841 + 724) * (518 + 967)}$$

$$= 109.63$$

$$\chi = 10.47$$

$$\text{Lower Limit} = 2.2(1 - 1.96/10.47)$$

$$= 1.876$$

$$\text{Upper Limit} = 2.2(1 + 1.96/10.47)$$

$$= 2.507$$

Case-Control Studies

- (b) Women who are current IUD users have a 2.2 times greater risk of PID than women who are not using any method of contraception. The 95% confidence interval means that the researcher was 95% confident that the true odds ratio is between 1.9 and 2.5. Another way of saying this is, we know that the OR is statistically larger than 1.0, since the lower confidence limit is greater than 1.0. Current IUD users are at greater risk of PID than women who are not currently using any method of contraception.
 - (c) Other questions the investigators might ask: Were cases and controls comparable? For example, perhaps controls should have been selected from a minor surgery clinic in the same hospital. Were other possible risk factors similar for cases and controls (e.g., number of sexual partners)?
 - (d) Age, marital status, education, history of sexually transmitted diseases, frequency of intercourse, previous episodes of PID, contraceptive history, previous episodes of IUD use, number of sexual partners.
2. Using Example 10.19, interpret the odds ratio.
- (a) The risk of developing cervical cancer is approximately four times higher among women who were sexually active at age 15 years or younger than among women who were sexually active later in life. Since the confidence interval does not include 1.0, then the hypothesis that there is no difference between women with cervical dysplasia and controls is rejected.
3. Using Example 10.20, analyze and interpret the data.
- (a) $OR = 0.5$ (95% CI: 0.4 - 0.7)
 - (b) The risk of developing PID for women currently using OCs is 0.5 (half) that of women not currently using OCs.

4. Interpret the odds ratio and confidence interval.
- (a) The risk of breast cancer among women who use DMPA is twice the risk among women who have never used DMPA. Since the confidence interval does not include 1.0, then the hypothesis that there is no difference in breast cancer risk between women who have used DMPA and women who have not used DMPA is rejected; the odds ratio is statistically significant.
5. True or false.
- (a) F Introduction. No, groups are identified according to having or not having a particular health problem.
- (b) T Introduction
- (c) T Introduction
- (d) F Control Selection. No, the control group provides an estimate of the rate of exposure if there was *no* association between the exposure and the health problem.
- (e) T Case Selection
- (f) T Control Selection
- (g) T Control Selection
- (h) F Control Selection. No, cases and controls should be matched on variables related to the exposure and the health problem.
- (i) F Control Selection. No, matching during data collection must be maintained for analysis.
- (j) F Control Selection. No, matching on relatives may be important for some investigations and will likely result in matching on other variables, such as diet and behaviors.
- (k) T Exposure Definition

Case-Control Studies

6. Multiple choice.

- 6.1 b **Exposure Definition.** Total months of condom use or duration of condom use is a time-related measure of exposure.
- 6.2 e **Exposure Definition.** Ever breast-fed is a frequency measure of exposure.
- (a) T **Selection Bias**
- (b) T **Selection Bias**
- (c) T **Information Bias**
- (d) T **Information Bias**
- (e) T **Information Bias**
- (f) T **Information Bias**
- (g) F **Information Bias.** No, excluding study participants with a poor memory is not a method for reducing bias. In fact, such an exclusion might introduce bias.

7. True or false.

- (a) T **Selection Bias**
- (b) T **Selection Bias**
- (c) T **Information Bias**
- (d) T **Information Bias**
- (e) T **Information Bias**
- (f) T **Information Bias**

- (g) F Information Bias. No, excluding study participants with a poor memory is not a method for reducing bias. In fact, such an exclusion might introduce bias.

8. Multiple choice.

8.1 a Data Collection Methods. We are interested in medical events that occurred prior to the diagnosis or enrollment in a case-control study.

8.2 c Data Analysis Methods

8.3 c Bias

9. True or false.

(a) T Advantages #1

(b) F Disadvantages #3. No, incidence rates cannot be computed from case-control data.

(c) F Advantages #2. No, case-control designs are ideal for studying health problems with long latency.

(d) T Advantages #3

(e) T Disadvantages #2

(f) F Disadvantages. No, incidence rates cannot be computed from case-control data.

(g) F Data Analysis Methods. No, the odds ratio is an appropriate estimate of the relative risk when the incidence of the health problem is low.

(h) T Data Analysis Methods

(i) F Data Analysis Methods. A variable may be both a confounder and an effect modifier.

(j) T Data Analysis Methods

Case-Control Studies

(k) T Data Analysis Methods

(l) T Data Analysis Methods

10. The answer given here is a suggested answer and is based on the report of the actual study referenced.

Example 10.23

Oral Contraceptive Use in the Past and Myocardial Infarction

Problem: Is long-term OC use in the past associated with the occurrence of myocardial infarction?

Research Hypothesis: Long-term OC use after discontinuation is associated with a myocardial infarction.

Study Design: Multicenter hospital-based case-control study

Cases: Women younger than 65 years old with an admission diagnosis of first myocardial infarction, defined according to World Health Organization criteria. Cases with a history of rheumatic valvular disease, cardiomyopathy, or cardiac surgery (including coronary artery bypass surgery) were excluded.

Controls: Women younger than 65 years old without a history of myocardial infarction, rheumatic valvular disease, cardiomyopathy, or cardiac surgery. Control women were admitted for nonmalignant and nongynecologic symptoms that were not related to OC use.

Exposure: The woman's report that she used medicines including oral contraceptives and noncontraceptive estrogens. The timing and brands for all medicines taken were recorded.

Data Collection Methods: Personal interview and medical record review.

Data Analysis:

<u>Myocardial Infarction</u>	<u>OC Use</u>		<u>Total</u>
	<u>Past Use</u>	<u>Never Use</u>	
Yes	291	613	904
No	673	1,047	1,720
Total	964	1,660	2,624

(Rosenberg et al., 1990)

11. Compute the odds ratios and confidence intervals.

(a) IUD OR = 0.4 (95% CI: 0.3 - 0.6)

Barrier or
natural OR = 0.3 (95% CI: 0.25 - 0.45)

OC OR = 0.1 (95% CI: 0.1 - 0.19)

(b) Women who are currently using IUDs, traditional methods of contraception, and OCs have a lower risk of an ectopic pregnancy than women who are currently not using any method. Since the confidence intervals do not include 1.0, the hypotheses of no association between ectopic pregnancy risk for women who use these methods compared with women who do not practice contraception are rejected.

(c) OR = 3.2 (95% CI: 2.1 - 5.0)

(d) Women who are currently using an IUD have a greater risk of an ectopic pregnancy than women currently using OCs. Since the confidence interval does not include 1.0, the hypothesis of no association between ectopic pregnancy risk for women currently wearing an IUD compared with women currently using OCs is rejected.

References

- Adolusi B. Carcinoma of the cervix uteri in Ibadan: coital characteristics. *Int J Gynaecol Obstet* 1977;15:5-11.
- Bennett AE, Ritchie K. *Questionnaires in medicine--a guide to their design and use*. London: Oxford University Press, 1975.
- The Boston Collaborative Drug Surveillance Program. Oral contraceptives and venous thromboembolic disease, surgically confirmed gallbladder disease and breast tumors. *Lancet* 1973;1:1399-1404.
- Burkman RT, the Women's Health Study. Association between intrauterine devices and pelvic inflammatory disease. *Obstet Gynecol* 1981;57:269-276.
- The Cancer and Steroid Hormone Study of the Centers for Disease Control, National Institute of Child Health and Human Development. Oral contraceptive use and the risk of breast cancer. *N Engl J Med* 1986;315:405-411.
- The Cancer and Steroid Hormone Study of the Centers for Disease Control, National Institute of Child Health and Human Development. The reduction in risk of ovarian cancer associated with oral contraceptive use. *N Engl J Med* 1987;316:650-655.
- Coulter A, Vessey M, McPherson K. The ability of women to recall their oral contraceptive histories. *Contraception* 1986;33:127-137.
- Grimes DA, Techman T. Legal abortion and placenta previa. *Am J Obstet Gynecol* 1984;149:501-504.
- Hogue CJR, Rubin GL, Schulz K. [Unpublished report]. Reproductive epidemiology in a nutshell. [1985]. Located at: Centers for Disease Control, National Center for Chronic Disease Prevention and Health Promotion, Atlanta, Georgia.
- Lee NC, Rosero-Bixby L, Oberle MW, Grimaldo C, Whatley AS, Rovira EZ. A case-control study of breast cancer and hormonal contraception in Costa Rica. *J Natl Cancer Inst* 1987;79:1247-1254.

- McPherson K, Neil A, Vessey MP, Doll R. Oral contraception and breast cancer. *Lancet* 1983;2:1414-1415.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22:719-748.
- Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol* 1976;103:226-235.
- Ory HW, the Women's Health Study. Ectopic pregnancy and intrauterine contraceptive devices: new perspectives. *Obstet Gynecol* 1981;57:137-144.
- Rosenberg L, Mitchell AA, Shapiro S, Slone D. Selected birth defects in relation to caffeine-containing beverages. *JAMA* 1982;247:1429-1432.
- Rosenberg L, Palmer JR, Lesko SM, Shapiro S. Oral contraceptive use and the risk of myocardial infarction. *Am J Epidemiol* 1990;131:1009-1016.
- Rubin GL, Ory HW, Layde PM. Oral contraceptives and pelvic inflammatory disease. *Am J Obstet Gynecol* 1982;144:630-635.
- Schlesselman JJ. Case-control studies. New York: Oxford University Press, 1982.
- Slone D, Shapiro S, Kaufman DW, Rosenberg L, Miettinen OS, Stolley PD. Risk of myocardial infarction in relation to current and discontinuous use of oral contraceptives. *N Engl J Med* 1981;305:420-424.
- Sudman S, Bradburn NM. Asking questions. San Francisco: Jossey-Bass Publishers, 1983.
- Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet* 1955;19:251-253.

