CHAPTER 8

Designing Case–Control Studies

Thomas B. Newman, Warren S. Browner, Steven R. Cummings, and Stephen B. Hulley

In Chapter 7 we introduced cohort studies, in which the sequence of the measurements is the same as the chronology of cause and effect: predictor variables are measured first, then outcomes are observed during follow-up. In contrast, in a **case–control** study the investigator works backward. She begins by choosing one sample of people with the outcome (the cases) and another sample of people without that outcome (the controls); she then compares the levels of predictor variables in the two samples to see which predictors are associated with the outcome. For example, a case–control study might involve assembling a group of cases of ocular melanoma and a sample of healthy controls, followed by gathering data from each group about previous exposure to arc welding to estimate how that exposure affects the risk of ocular melanoma. The case–control design is relatively **inexpensive** and uniquely **efficient** for studying **rare diseases**.

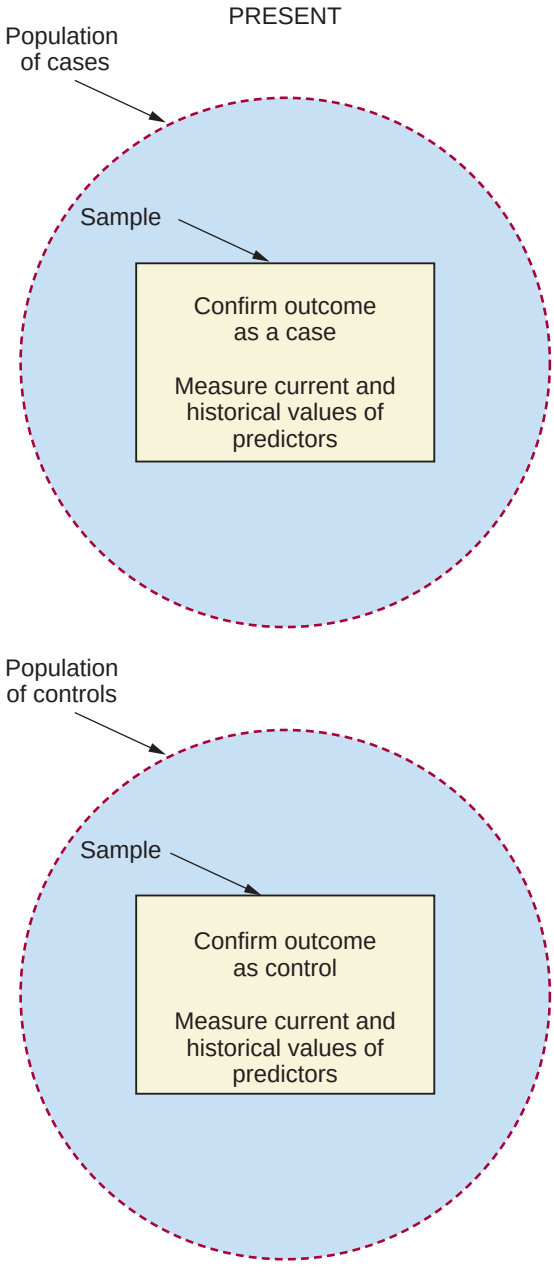
This chapter also presents several variations on the simple case–control design noted above. A ***nested* case–control** design compares the incident cases nested in a cohort study with controls drawn at random from the rest of the cohort; this design controls sampling and measurement bias and saves money if the predictors are expensive measurements that can be made on stored specimens or images collected at the outset of the cohort study. An ***incidence-density* case–control** design allows investigators to analyze risk relationships, taking into account changes over time in risk factor levels and loss to follow-up. And a **nested case–cohort** design allows a random sample of the entire cohort to serve as the control for several different sets of cases. The chapter ends with advice on choosing among the observational study designs discussed in Chapters 7 and 8.

CASE–CONTROL STUDIES

Because most diseases are relatively uncommon, both cohort and cross-sectional studies of general population samples are expensive designs, requiring thousands of subjects to identify risk factors for a rare disease like stomach cancer. As noted in Chapter 7, a **case series** of patients with the disease can identify an obvious risk factor (such as injection drug use for AIDS), using prior knowledge of the prevalence of the risk factor in the general population. For most risk factors, however, it is necessary to assemble a reference group, so that exposure to the risk factor in subjects with the disease (cases) can be compared with exposure to the risk factor among subjects without the disease (controls).

Case–control studies are **retrospective** (Figure 8.1). The study identifies one group of subjects with the disease and another without it, then looks backward to find differences in predictor variables that may explain why the cases got the disease and the controls did not (Example 8.1).

Case–control studies began as epidemiologic studies to identify risk factors for diseases. For this reason, and because it makes the discussion easier to follow, we generally refer to "cases" as those with the disease. However, the case–control design can also be used to look at other uncommon outcomes, such as disability among those who already have a disease. In addition, when



■ **FIGURE 8.1** In a case–control study, the steps are to:

- Define selection criteria and recruit one sample from a population of cases and a second sample from a population of controls.
- Measure current values of relevant variables, often supplemented by historical information.

undesired outcomes are the rule rather than the exception, the cases in a case–control study may be the rare patients who have had a good outcome, such as recovery from a usually fatal disease.

Case–control studies are the "house red" on the research design wine list: more modest and a little riskier than the other selections, but much less expensive and sometimes surprisingly good. The design of a case–control study is challenging because of the increased opportunities for bias, but there are many examples of well-designed case–control studies that have yielded important results. These include the links between maternal diethylstilbestrol use and vaginal cancer in daughters (a classic study that provided a definitive conclusion based on just seven cases!) (1), and between prone sleeping position and sudden infant death syndrome (2), a simple result that has saved thousands of lives (3).

EXAMPLE 8.1 Case–Control Study

Because intramuscular vitamin K is given routinely to newborns in the United States, a pair of studies reporting a doubling in the risk of childhood cancer among those who had received intramuscular vitamin K caused quite a stir (4, 5). To investigate this association further, German investigators (6)

1. **Selected the sample of cases.** 107 children with leukemia from the German Childhood Cancer Registry.
2. **Selected the sample of controls.** 107 children matched by sex and date of birth and randomly selected from children living in the same town as the case at the time of diagnosis (from local government residential registration records).
3. **Measured the predictor variable.** Reviewed medical records to determine which cases and controls had received intramuscular vitamin K in the newborn period.

The authors found 69 of 107 cases (64%) and 63 of 107 controls (59%) had been treated with vitamin K, for an odds ratio of 1.3 (95% confidence interval [CI], 0.7 to 2.3). (See Appendix 8A for the calculation.) Therefore, this study did not confirm the existence of an association between the receipt of vitamin K as a newborn and subsequent childhood leukemia. The point estimate and upper limit of the 95% CI leave open the possibility of a clinically important increase in leukemia in the population from which the samples were drawn, but several other studies, and an analysis using an additional control group in the cited study, also failed to confirm the association (7, 8).

Case–control studies cannot yield estimates of the incidence or prevalence of a disease because the proportion of study subjects who have the disease is determined by how many cases and how many controls the investigator chooses to sample, rather than by their proportions in the population. Case–control studies do provide descriptive information on the characteristics of the cases and, more important, an estimate of the strength of the association between each predictor variable and the outcome. These estimates are in the form of odds ratios, which approximate the relative risk if the risk of the disease in both exposed and unexposed subjects is relatively low (about 10% or less; see Appendix 8B).

Strengths of Case–Control Studies

Efficiency for Rare Outcomes

One of the major strengths of case–control studies is their rapid, high yield of information from relatively few subjects. Consider a study of the effect of circumcision on subsequent carcinoma of the penis. This cancer is very rare in circumcised men but is also rare in uncircumcised men, whose lifetime cumulative incidence is about 0.16% (9). To do a cohort study with a reasonable chance (80%) of detecting even a very strong risk factor (say a relative risk of 50) would require following more than 6,000 men for many years, assuming that roughly equal proportions were circumcised and uncircumcised. A randomized clinical trial of circumcision at birth would require the same sample size, but the cases would occur at a median of 67 years after entry into the study—it would take three generations of investigators to follow the subjects!

Now consider a case–control study of the same question. For the same chance of detecting the same relative risk, only 16 cases and 16 controls (and not much time or effort from the investigators) would be required. For diseases that are either rare or have long latent periods between exposure and disease, case–control studies are not only far more efficient than other designs, they are often the only feasible option.

Usefulness for Generating Hypotheses

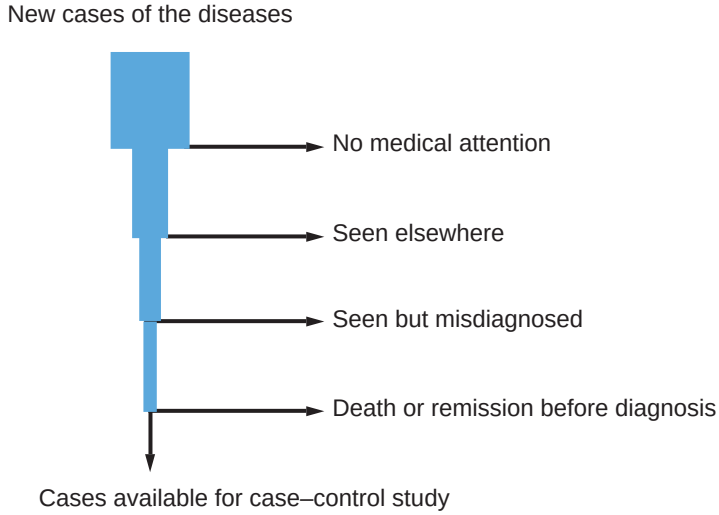
The retrospective approach of case–control studies, and their ability to examine a large number of predictor variables, makes them useful for generating hypotheses about the causes of a new outbreak of disease. For example, a case–control study of an epidemic of deaths from acute renal failure in Haitian children found an odds ratio of 53 for ingestion of locally manufactured acetaminophen syrup. Further investigation revealed that the renal failure was due to poisoning by diethylene glycol, which was found to contaminate the acetaminophen syrup (10), a problem that unfortunately has recurred (11).

Weaknesses of Case–Control Studies

Case–control studies have great strengths, but they also have major disadvantages. First, only one outcome can be studied (the presence or absence of the disease that was the criterion for drawing the two samples), whereas cohort and cross-sectional studies (and clinical trials) can study several outcome variables. Second, as mentioned, the information available in case–control studies is limited: There is no direct way to estimate the incidence or prevalence of the disease, nor the attributable or excess risk, unless the investigator also knows the exact population and time period from which the cases arose. But the biggest weakness of case–control studies is their **susceptibility to bias**. This bias comes chiefly from two sources: the **separate sampling** of the cases and controls, and the **retrospective measurement** of the predictor variables. These two problems and the strategies for dealing with them are the topic of the next two sections.

Sampling Bias and How to Control It

The sampling in a case–control study begins with the cases. Ideally, the sample of cases would include everyone who developed the disease under study, or a random selection from those cases. An immediate problem comes up, however: How do we know who has developed the disease and who has not? In cross-sectional and cohort studies the disease is systematically sought in all the study participants, but in case–control studies the cases must be sampled from patients in whom the disease has already been diagnosed and who are available for study. This sample may not be representative of all patients who develop the disease because those who are undiagnosed, misdiagnosed, unavailable for study, or dead are unlikely to be included (Figure 8.2).



■ **FIGURE 8.2** Some reasons that the cases in a case–control study may not be representative of all cases of the disease.

In general, sampling bias matters when the sample of cases is unrepresentative with respect to the risk factor being studied. Diseases that almost always require hospitalization and are straightforward to diagnose, such as hip fracture and traumatic amputation, can be sampled safely from diagnosed and accessible cases, at least in developed countries. On the other hand, conditions that may not come to medical attention are more difficult to study with case–control studies because of the selection that precedes diagnosis. For example, women seen in a gynecologic clinic with first-trimester spontaneous abortions would probably differ from the entire population of women experiencing spontaneous abortions, many of whom do not seek medical attention. Thus women with a prior history of infertility would be over-represented in a clinic-based sample, while those with poor access to prenatal care would be under-represented. If a predictor variable of interest is associated with gynecologic care in the population (such as past use of an intrauterine device [IUD]), sampling cases from the clinic could be an important source of bias. If, on the other hand, a predictor is unrelated to gynecologic care (such as blood type), there would be less likelihood of a clinic-based sample being unrepresentative.

Although it is important to think about these issues, the selection of cases is often limited to the accessible sources of subjects. The sample of cases may not be entirely representative, but it may be all that the investigator has to work with. The difficult decisions faced by an investigator designing a case–control study then relate to the more open-ended task of selecting appropriate controls. The general goal is to sample controls from the population who would have become a case in the study if they had developed the disease. Four strategies for sampling controls follow:

- **Clinic- or hospital-based controls.** One strategy to compensate for the possible selection bias caused by obtaining cases from a clinic or hospital is to select controls from the same facility or facilities. For example, in a study of past use of an IUD as a risk factor for spontaneous abortion, controls could be sampled from a population of women seeking care for other problems (e.g., vaginitis) at the same gynecologic clinic. Compared with a random sample of women from the same area, these controls would presumably better represent the population of women who, if they had a spontaneous abortion, would have come to the clinic and become a case.

  However, selection of an unrepresentative sample of controls to compensate for an unrepresentative sample of cases can be problematic. If the risk factor of interest causes a medical problem for which the controls seek care, the prevalence of the risk factor in the control group will be falsely high, diminishing or reversing the association between the risk factor and the outcome. If, for example, many women in the control group sought attention at the clinic for a medical condition associated with past use of an IUD (e.g., infertility from older models of IUDs), there would be an excess of former IUD users among the controls, reducing the size of the association between past IUD use and spontaneous abortion in the study.

  Because hospital- and clinic-based control subjects often have conditions that are associated with the risk factor(s) being studied, these types of controls can produce misleading findings. Thus it is essential to consider whether the convenience of using hospital- or clinic-based controls is worth the possible threat to the validity of the study.
- **Using population-based samples of cases and controls.** Because of the rapid increase in the use of disease registries in geographic populations and within health plans, population-based case–control studies are now possible for many diseases. Cases obtained from such registries are generally representative of the general population of patients in the area with the disease, thus simplifying the choice of a control group: It should be a representative sample of "non-cases" from the population covered by the registry. In Example 8.1, all residents of the town were registered with the local government, making selection of such a sample straightforward.

  When registries are available, population-based case–control studies are the most desirable design. As a disease registry approaches completeness and the population it covers approaches

stability (no migration in or out), a population-based case–control study approaches a case–control study that is nested within a cohort study or clinical trial (page 104) assuming that the controls can be identified and enrolled. Those latter tasks are relatively straightforward when the population has been enumerated and these records are available to investigators, as in the vitamin K and leukemia study described in Example 8.1. Lacking such registration records, a commonly used approach is random digit dialing of (landline) phone numbers with prefixes in the region covered by the registry. (When controls are selected this way, the cases who have no landline telephone need to be excluded.) With increasing numbers of households with mobile phones only, this approach has become problematic (12). Random-digit dialing including cell phone numbers is possible, but must be done carefully, immediately ending the call if the recipient is driving and avoiding calls for which the recipient might be charged (13).

It's important to recognize, however, that bias can be introduced any time subjects need to be contacted to obtain information because some subjects (say, those who do not speak English, or who are hard of hearing) may be less likely to be included. A similar problem can occur any time informed consent is needed.

- **Using two or more control groups.** Because selection of a control group can be so tricky, particularly when the cases may not be a representative sample of those with disease, it is sometimes advisable to use two or more control groups selected in different ways. The Public Health Service study of Reye's syndrome and medications (14), for example, used four types of controls: emergency room controls (seen in the same emergency room as the case), inpatient controls (admitted to the same hospital as the case), school controls (attending the same school or day care center as the case), and community controls (identified by random-digit dialing). The odds ratios for salicylate use in cases compared with each of these control groups were all at least 30 and highly statistically significant. The consistent finding of a strong association using control groups that would have different sampling biases strengthens the inference that there is a real association in the population.

  Unfortunately, few associations have odds ratios anywhere near that large, and the biases associated with different strategies for selecting controls may cause the results using different control groups to conflict with one another, thereby revealing the inherent fragility of the case–control design for the research question at hand. When this happens, the investigator should seek additional information (e.g., the chief complaint of clinic-based controls) to try to determine the magnitude of potential biases from each of the control groups (Chapter 9). In any case it is better to have inconsistent results and conclude that the answer is not known than to have just one control group and draw the wrong conclusion.
- **Matching.** Matching is a simple method of ensuring that cases and controls are comparable with respect to major factors that are related to the disease but not of interest to the investigator. So many risk factors and diseases are related to age and sex, for example, that the study results may be unconvincing unless the cases and controls are comparable with regard to these two variables. One approach to avoiding this problem is to choose controls that match the cases on these constitutional predictor variables. However, matching does have substantial disadvantages, particularly if modifiable predictors such as income or serum cholesterol level are matched. The reasons for this and the alternatives that are often preferable to matching are discussed in Chapter 9.

Differential Measurement Bias and How to Control It

The second major weakness of case–control studies is the risk of bias due to **measurement error**. This is caused by the retrospective approach to measuring the predictor variables: both cases and control may be asked to recall exposures that happened years before. Unfortunately, people's memories for past exposures are imperfect. If they are similarly imperfect in cases and

controls, the problem is called **nondifferential misclassification** of the exposure, which makes it more difficult to find associations. (In epidemiologic terms, the odds ratio is biased toward 1.) Of greater concern, however, being diagnosed with a disease may lead cases to remember or report their exposures differently from controls; this **differential misclassification** of exposure, called **recall bias**, has unpredictable effects on associations measured in a study.

For example, widespread publicity about the relationship between sun exposure and malignant melanoma might lead cases diagnosed with that cancer to recall their history of sun exposure differently from controls. Cockburn et al. (15) found some evidence of this in a clever study of twins discordant for melanoma: The matched odds ratio for sunbathing as a child was 2.2 (95% CI 1.0 to 4.7) when the twin with melanoma was asked which twin had sunbathed more as a child, but only 0.8 (0.4 to 1.8) when the co-twin without melanoma was asked the same question. However, for some other questions, such as which twin tanned or burned more easily, there was no evidence of recall bias.

Recall bias cannot occur in a cohort study because the subjects are asked about exposures before the disease has been diagnosed. A case–control study of malignant melanoma nested within a cohort with sun exposure data collected years earlier provided a direct test of recall bias: The investigators compared self-reported sun exposure in cases and controls both before and after the case was diagnosed with melanoma (16). The investigators found some inaccuracies in recollections of exposure in both cases and controls, but little evidence of recall bias (16). Thus, while it is important to consider the possibility of recall bias, it is not inevitable (17).

In addition to the strategies set out in Chapter 4 for controlling bias in measurements (standardizing the operational definitions of variables, choosing objective approaches, supplementing key variables with data from several sources, etc.), here are two specific strategies for avoiding bias in measuring exposures in case–control studies:

- **Use data recorded before the outcome occurred.** It may be possible, for example, to examine perinatal medical records in a case–control study of intramuscular vitamin K as a risk factor for cancer. This excellent strategy is limited to the extent that recorded information about the risk factor of interest is available and reliable. For example, information about vitamin K administration was often missing from medical records, and how that missing information was treated affected results of some studies of vitamin K and subsequent cancer risk (8).
- **Use blinding.** The general approach to blinding was discussed in Chapter 4, but there are some issues that are specific to designing interviews in case–control studies. In theory, both observers and study subjects could be blinded to the case–control status of each subject and to the risk factor being studied; thus, four types of blinding are possible (Table 8.1).

TABLE 8.1 APPROACHES TO BLINDING IN A CASE–CONTROL STUDY

PERSON BLINDED	BLINDING CASE–CONTROL STATUS	BLINDING RISK FACTOR MEASUREMENT
Subject	Possible if both cases and controls have diseases that could plausibly be related to the risk factor	Include "dummy" risk factors and be suspicious if they differ between cases and controls May not work if the risk factor for the disease has already been publicized
Observer	Possible if cases are not externally distinguishable from controls, but subtle signs and statements volunteered by the subjects may make it difficult	Possible if interviewer is not the investigator, but may be difficult to maintain

Ideally, neither the study subjects nor the observers should know which subjects are cases and which are controls. In practice, this is often difficult. The subjects know whether they are sick or well, so they can be blinded to case–control status only if controls are also ill with diseases that they believe might be related to the risk factors being studied. Efforts to blind interviewers are hampered by the obvious nature of some diseases (an interviewer can hardly help noticing if the subject is jaundiced or has had a laryngectomy), and by the clues that interviewers may discern in the subject's responses.

Blinding to specific risk factors being studied is usually easier than blinding to case–control status. Case–control studies are often the first step in investigating an illness, so there may not be just one risk factor of particular interest. Thus, the study subjects and the interviewer can be kept in the dark about the study hypotheses by including "dummy" questions about plausible risk factors not associated with the disease. For example, in a study of honey consumption as a risk factor for infant botulism, equally detailed questions about yogurt and bananas could be included in the interview. This type of blinding does not prevent differential bias, but it allows an estimate of whether it is a problem: If the cases report more exposure to honey but no increase in the other foods, then differential measurement bias is less likely. This strategy would not work if the association between eating honey and infant botulism had previously been widely publicized, or if some of the dummy risk factors turned out to be real ones.

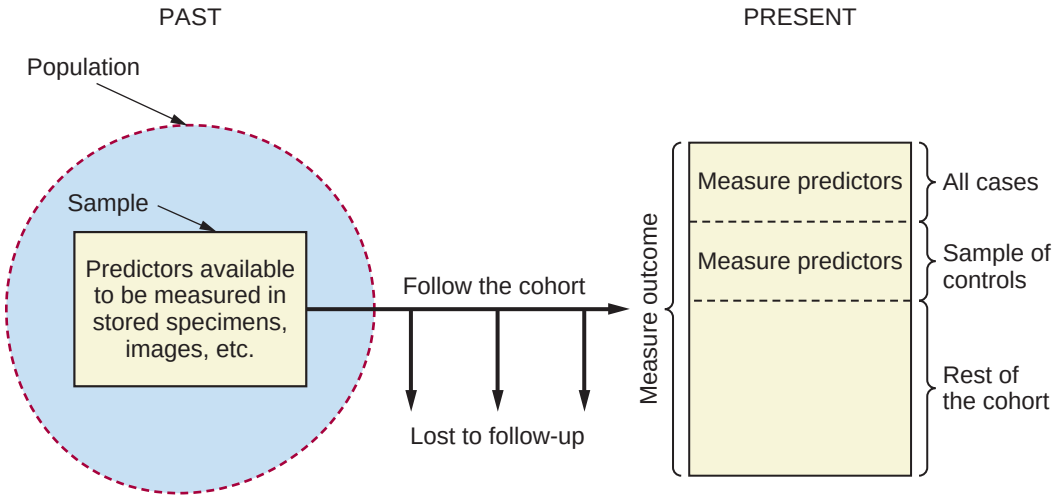
Blinding the observer to the case–control status of the study subject is a particularly good strategy for **laboratory measurements** such as blood tests and x-rays. Blinding under these circumstances is easy and should always be done, simply by having someone other than the individual who will make the measurement apply a coded identification label to each specimen (or patient). The importance of blinding was illustrated by 15 case–control studies comparing measurements of bone mass between hip fracture patients and controls; much larger differences were found in the studies that used unblinded measurements than in the blinded studies (18).

NESTED CASE–CONTROL, INCIDENCE-DENSITY NESTED CASE–CONTROL, AND CASE–COHORT STUDIES

A **nested case–control** design has a case–control study "nested" within a defined cohort (Figure 8.3). The cohort may already have been defined by the investigator as part of a formal cohort study, often including banking of specimens, images, and so on, to be analyzed in the future after outcomes occur. Alternatively, the investigator can design a nested case–control study *de novo*, in a cohort that is not already defined, in which case defining the cohort will be the first step.

The investigator begins by identifying a cohort of subjects at risk for the outcome that is large enough to yield sufficient numbers of cases to answer the research question, and that provides the ability to measure the exposure variable, either because specimens have been banked or medical records (or subjects) with exposure information are available. As described in Chapter 7, definition of the cohort will include the specific inclusion and exclusion criteria that define a population at risk. In addition, the **date of entry** into the cohort must be clear for each subject. This could be a fixed date (e.g., everyone meeting inclusion criteria who was enrolled in a health plan on January 1, 2008), or it could be a variable date on which a period at risk begins (e.g., the date of enrollment in a cohort study or the date of first myocardial infarction in a study of risk factors for recurrent myocardial infarction).

The investigator next describes the criteria that define the occurrence of the outcome of interest, which in all cases will be after the date of entry into the cohort and before the end of the defined follow-up period. If the outcome is rare, follow-up close to complete, and a single measurement of the exposure at baseline is sufficient, then it is simple. The investigator identifies all the individuals in the cohort who developed the outcome by the end of follow-up (the cases) and then selects a random sample of the subjects who were also part of the cohort but did not develop the outcome (the controls). The investigator then measures the predictor variables



■ **FIGURE 8.3** A *nested* case–control study can be either prospective or retrospective. For the retrospective version, the steps are to

- Identify a cohort from the population with previously stored specimens, images, and other data.
- Measure the outcome variable that distinguishes cases from controls.
- Measures predictor variables in specimens, images, and other data stored since the cohort was formed, as well as other variables, in all the cases and in a sample of the non-cases (controls).

for cases and controls, and compares levels of the risk factor in cases to the levels in the sample of controls. This is a simple nested case–control study (Example 8.2).

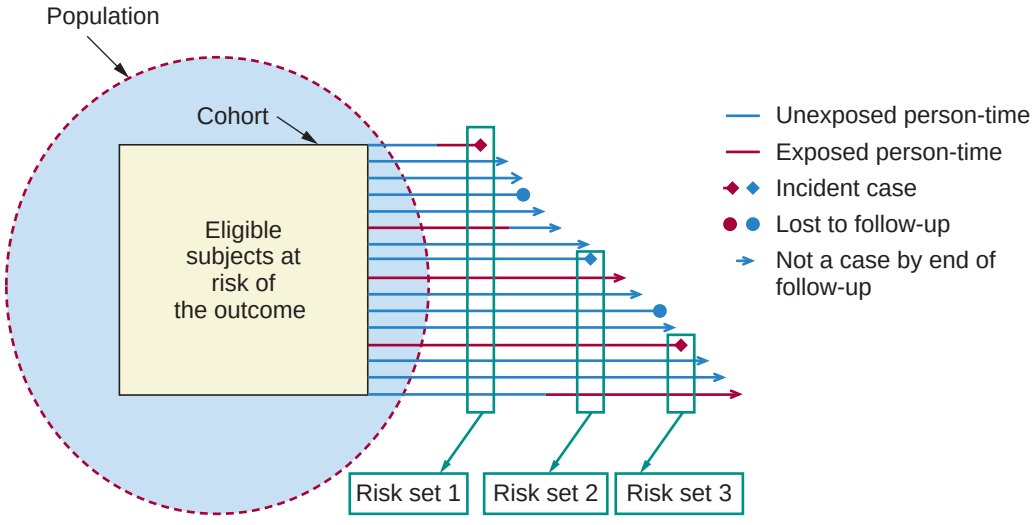
If follow-up is variable or **incomplete**, or the exposure of interest **varies over time**, a single measurement of exposure at entry into the cohort in the cases and a random sample of controls

EXAMPLE 8.2 Simple Nested Case–Control Design

To determine whether higher levels of sex hormones increased the risk of breast cancer, Cauley (19) and colleagues conducted a nested case–control study. The basic steps in performing this study were to:

1. **Identify a cohort.** The investigators used the Study of Osteoporotic Fractures (SOF) cohort. This was a good choice because serum samples of members of this cohort had been drawn by the same investigators during the baseline examination and put into frozen storage at –190°C with the expectation that just such a study would be designed.
2. **Identify cases at the end of follow-up.** Based on responses to follow-up questionnaires and review of death certificates, the investigators identified 97 subjects who had developed a first occurrence of breast cancer during 3.2 years of follow-up.
3. **Select controls.** The investigators selected a random sample of 244 women in the cohort who did not develop breast cancer during that follow-up period.
4. **Measure predictors.** Levels of sex hormones, including estradiol and testosterone, were measured in the samples of frozen serum from the baseline examination of cases and controls. The laboratory was blinded to whether the samples came from cases or controls.

Women who had high levels of either estradiol or testosterone had a threefold increase in the risk of a subsequent diagnosis of breast cancer compared with women who had very low levels of these hormones.



■ **FIGURE 8.4** An *incidence-density* nested case–control study can be either prospective or retrospective. For the prospective version, the steps are to:

- Define selection criteria and recruit a cohort from the population.
- Define the date of entry for each member of the cohort to align follow-up times.
- Store specimens, images, etc for later analysis.
- Follow the cohort to identify cases and the date they were diagnosed.
- Sample one or more controls for each case from "risk sets," defined as members of the cohort who have been followed for the same amount of time as the case and have not become a case, died, or been lost to follow-up at the time the case was diagnosed.
- Measure predictor variables in specimens, images, etc. stored since baseline, as well as other current variables, on cases and matched controls.

will not be sufficient. In that case it is better to design an ***incidence-density* nested case–control study** and sample the controls from **risk sets**, defined for each case as it occurs as the members of the cohort who were followed the same length of time as the case but had not yet become cases (Figure 8.4). As is the case for any other form of matching of controls to cases, this matching on follow-up time needs to be accounted for in the analysis.

For example, if entry in the cohort was a fixed date (e.g., January 1, 2008), the controls for a case diagnosed on July 1, 2009, would be sampled from among the subjects who had not yet developed the outcome as of July 1, 2009. If the date of entry into the cohort was variable, controls for a case diagnosed 18 months after entry would be sampled from among those who had not yet become a case after 18 months of follow-up. Depending on the research hypothesis of the investigator, values of the exposure at entry or at some point after entry could be compared between cases and controls.

This sampling according to risk sets introduces the complexity that the same subject may be selected as a control for a case that occurs early in follow-up and later become a case himself, perhaps after his value for his exposure variable changes. In effect, what this design does (with the help of appropriate statistical analysis) is sequentially consider chunks of person-time at risk, for each chunk using values of predictor variables to predict occurrence of cases in that chunk of person-time, with the boundaries of each chunk defined by the occurrence of the cases. This is called an **incidence-density** design (Example 8.3).

A **nested case–*cohort*** design is similar to the simple nested case–control design except that, instead of selecting controls who did not develop the outcome of interest, the investigator selects a random sample of all the members of the cohort, regardless of outcomes. A few subjects who are part of that random sample may have developed the outcome (the number is very small when the outcome is uncommon). An advantage of the case–cohort design is that a

single random sample of the cohort can provide the controls for several case–control studies of different outcomes. In addition, the random sample of the cohort provides information on the overall prevalence of risk factors in the cohort.

EXAMPLE 8.3 "Incidence-Density" Nested Case–Control Design

To investigate a possible association between the oral antidiabetes drug pioglitazone (Actos®) and bladder cancer, investigators from Montreal (20) performed a case–control study nested within the United Kingdom General Practice Research Database, which contains complete primary care medical records for more than 10 million people enrolled in more than 600 general practices in the UK. The steps were:

1. **Identify the cohort and time period at risk.** The investigators included adults with their first ever prescription for an oral antidiabetes drug between January 1, 1988, and December 31, 2009, who had been followed in the database for at least 1 year before that prescription and who were at least 40 years old at the time of that prescription. The date of this first antidiabetes drug prescription was the date of entry into the cohort. Participants were followed until a diagnosis of bladder cancer, death from any cause, end of registration with the general practice, or end of the study period on December 31, 2009, whichever came first. Subjects with a previous history of bladder cancer were excluded.
2. **Identify the cases, including dates of occurrence.** The investigators identified incident cases of bladder cancer using "Read codes" (a system for coding diagnoses validated in the general practice research database [21]). To account for the expectation that the effect of pioglitazone on cancer risk would not be expected to be immediate, they excluded cases occurring in the first year after cohort entry. They identified 376 remaining bladder cancer cases.
3. **Sample controls from "risk sets" matched to each case.** The investigators sampled up to 20 controls for each case, matched on year of birth, year of cohort entry, sex, and duration of follow-up, who had not been diagnosed with bladder cancer up to the date of diagnosis of the case. The total number of matched controls was 6,699 (average number of controls per case = 17.8).[1]
4. **Define and measure predictors.** The primary predictor of interest was receipt of a prescription of either pioglitazone or rosiglitazone, another antidiabetes drug in the same class as pioglitazone. The prescription needed to be at least 1 year before the date of diagnosis of the case in the risk set. Four exposure levels were defined: prescription for pioglitazone only, rosiglitazone only, both, or neither.

The authors (appropriately) used conditional logistic regression to analyze the data; this accounts for the matched nature of the data and, because of the risk-set sampling, allows estimation of adjusted rate ratios (22). They found adjusted rate ratios of 1.83 (95% CI 1.10 to 3.05) for exclusive pioglitazone use, 1.14 (95% CI 0.78 to 1.68) for exclusive rosiglitazone use, and 0.78 (95% CI 0.18 to 3.29) for use of both. (The wide confidence interval on the last group reflects a much smaller sample size [$N = 2$ cases and 56 controls]). They also found evidence of dose-response relationship between pioglitazone use and bladder cancer: The adjusted rate ratio for cumulative dose of 28 grams or more was 2.54 (1.05–6.14), P for dose-response trend = 0.03.

[1]We will point out in Chapter 9 that the gain in power from sampling more than four controls per case is slight, but in this case the additional cost was low because electronic data were already available. Even with 20 controls per case the nested case–control approach is much more computationally efficient than a retrospective cohort study.

Strengths

Nested case–control and case–cohort studies are especially useful for costly measurements on serum and other specimens or images that have been archived at the beginning of the study and preserved for later analysis. Making expensive measurements on all the cases and a sample of the controls is much less costly than making the measurements on the entire cohort.

This design preserves all the advantages of cohort studies that result from collecting predictor variables before the outcomes have happened. In addition, it avoids the potential biases of conventional case–control studies that cannot make measurements on fatal cases and that draw cases and controls from different populations.

Weaknesses

These designs share certain disadvantages of other observational designs: the possibilities that observed associations are due to the effect of unmeasured or imprecisely measured confounding variables and that baseline measurements may be affected by silent preclinical disease.

Other Considerations

Nested case–control and case–cohort designs have been used less often than they should be. An investigator planning large prospective studies should consider preserving biologic samples (e.g., banks of frozen sera) or storing images or records that are expensive to analyze for subsequent nested case–control analyses. She should ensure that the conditions of storage will preserve substances of interest for many years. It may also be useful to collect new samples or information during the follow-up period, which can also be used in the case–control comparisons.

CASE-CROSSOVER STUDIES

The case-crossover design is a variant of the case–control design that is useful for studying the short-term effects of intermittent exposures. As with ordinary case–control studies, these retrospective studies begin with a group of cases: people who have had the outcome of interest. However, unlike traditional case–control studies, in which the exposures of the cases are compared with exposures of a group of controls, in case-crossover studies each case serves as her own control. Exposures of the cases at the time (or right before) the outcome occurred are compared with exposures of those same cases at one or more other points in time.

For example, McEvoy et al. (23) studied cases who were injured in car crashes and reported owning or using a mobile phone. Using phone company records, they compared mobile phone usage in the 10 minutes before the crash with usage when the subjects were driving at the same time of day 24 hours, 72 hours, and 7 days before the crash. They found that mobile phone usage was more likely in the 10 minutes before a crash than in the comparison time periods, with an odds ratio of about 4. The analysis of a case-crossover study is like that of a matched case–control study, only the control exposures are exposures of the case at different time periods, rather than exposures of the matched controls. This is illustrated in Appendix 8A, scenario number 4. Case-crossover designs have been used in large populations to study time-varying exposures like levels of air pollution; associations have been found with myocardial infarction (24, 25), emergency room visits for respiratory disease (26), and even infant mortality (27).

CHOOSING AMONG OBSERVATIONAL DESIGNS

The pros and cons of the main observational designs presented in the last two chapters are summarized in Table 8.2. We have already described these issues in detail and will make only one final point here. Among all these designs, none is best and none is worst; each has its place and purpose, depending on the research question and the circumstances.

TABLE 8.2 ADVANTAGES AND DISADVANTAGES OF THE MAJOR OBSERVATIONAL DESIGNS

DESIGN	ADVANTAGES	DISADVANTAGES*
	Cross-sectional	
	Relatively short duration A good first step for a cohort study or clinical trial Yields prevalence of multiple predictors and outcomes	Does not establish sequence of events Not feasible for rare predictors or rare outcomes Does not yield incidence
	Cohort Designs	
All	Establishes sequence of events Multiple predictors and outcomes Number of outcome events grows over time Yields incidence, relative risk, excess risk	Often requires large sample sizes Less feasible for rare outcomes
Prospective cohort	More control over subject selection and measurements Avoids bias in measuring predictors	Follow-up can be lengthy Often expensive
Retrospective cohort	Follow-up is in the past Relatively inexpensive	Less control over subject selection and measurements
Multiple cohort	Useful when distinct cohorts have different or rare exposures	Bias and confounding from sampling distinct populations
	Case–Control	
	Useful for rare outcomes Short duration, small sample size Relatively inexpensive	Bias and confounding from sampling two populations Differential measurement bias Limited to one outcome variable Sequence of events may be unclear Does not yield prevalence, incidence, or excess risk unless nested within a cohort
	Hybrid Designs	
Nested case–control	Advantages of a retrospective cohort design, and less costly if measurement of predictors is expensive	Measurements of risk factors subject to bias if not previously measured or based on banked specimens or images stored previously; usually requires a preexisting defined cohort
Incidence-density nested case–control	Allows investigators to analyze risk relationships taking into account changes over time in risk factor levels and loss to follow-up	Requires measurements of risk factor levels and incidence of cases over time during follow-up; usually requires a preexisting defined cohort
Nested case–cohort	Same as nested case–control and can use a single control group for multiple case–control studies with different outcomes	Same as nested case–control
Case-crossover	Cases serve as their own controls, reducing random error and confounding	Requires that the exposure have only immediate, short-term effects

*All these observational designs have the disadvantage (compared with randomized trials) of being susceptible to the influence of confounding variables—see Chapter 9.

SUMMARY

1. In a **case–control study**, the prevalence of a risk factor in a sample of subjects who have the outcome of interest (**the cases**) is compared with the prevalence in a sample that does not (**the controls**). This design, in which people with and without the disease are sampled separately, is relatively **inexpensive** and uniquely **efficient** for studying **rare diseases**.
2. One problem with case–control studies is their susceptibility to **sampling bias**. Four approaches to reducing sampling bias are (a) to sample controls and cases in the **same** (admittedly unrepresentative) **way**; (b) to do a **population-based** study; (c) to use **several** control groups, sampled in different ways; and (d) to **match** the cases and controls.
3. The other major problem with case–control studies is their retrospective design, which makes them susceptible to **measurement bias** affecting cases and controls differentially. Such bias can be reduced by using **measurements of the predictor made prior to the outcome** and by **blinding** the subjects and observers.
4. The best way to **avoid both sampling and measurement bias** is to design a ***nested* case–control study** in which random samples of cases and controls are drawn from a larger cohort study at its conclusion. In addition to controlling both of these biases, expensive baseline measurements on serum, images, and so on, can be made at the end of the study on a relatively **small number of study subjects**.
5. The ***incidence-density* case–control design** allows investigators to analyze risk relationships, taking into account **changes over time** in **risk factor** levels and in the **availability of follow-up**.
6. The **nested case–*cohort*** design uses a random sample of the entire cohort in place of the non-cases; this can serve as a control group for studying **more than one outcome**, and provides direct information on the overall prevalence of risk factors in the cohort.
7. **Case-crossover studies** are a variation on the matched case–control design in which observations at two or more points in time allow each case to serve as her own control.

APPENDIX 8A
Calculating Measures of Association

1. **Cross-sectional study.** Reijneveld (28) did a cross-sectional study of maternal smoking as a risk factor for infant colic. Partial results are shown below:

TABLE 8A.1

	OUTCOME VARIABLE		
PREDICTOR VARIABLE	INFANT COLIC	NO INFANT COLIC	TOTAL
Mother smokes 15 to 50 cigarettes/day	15 *(a)*	167 *(b)*	182 *(a + b)*
Mother does not smoke	111 *(c)*	2,477 *(d)*	2,588 *(c + d)*
Total	126 *(a + c)*	2,644 *(b + d)*	2,770 *(a + b + c + d)*

Prevalence of colic with smoking mothers = $a/(a + b)$ = 15/182 = 8.2%.
Prevalence of colic with nonsmoking mothers = $c/(c + d)$ = 111/2,588 = 4.3%.
Prevalence of colic overall = $(a + c)/(a + b + c + d)$ = 126/2,770 = 4.5%.

$$\text{Relative prevalence}^2 = \frac{8.2\%}{4.3\%} = 1.9$$

$$\text{Excess prevalence}^2 = 8.2\% - 4.3\% = 3.9\%$$

In other words, colic was almost twice (1.9 times) as common, and occurred almost 4% more often, among children of smoking mothers.

2. **Case–control study.** The research question for Example 8.1was whether there is an association between intramuscular vitamin K and risk of childhood leukemia. The findings were that 69/107 leukemia cases and 63/107 controls had received vitamin K. A 2 × 2 table of these findings is as follows:

TABLE 8A.2

	OUTCOME VARIABLE: DIAGNOSIS	
PREDICTOR VARIABLE: MEDICATION HISTORY	CHILDHOOD LEUKEMIA	CONTROL
Intramuscular vitamin K	69(*a*)	63(*b*)
No intramuscular vitamin K	38(*c*)	44(*d*)
Total	107	107

$$\text{Relative risk} \approx \textit{odds ratio} = \frac{ad}{bc} = \frac{69 \times 44}{63 \times 38} = 1.27$$

[2]Relative prevalence and excess prevalence are the cross-sectional analogs of relative risk and excess risk.

Because the disease (leukemia in this instance) is rare, the odds ratio provides a good estimate of the relative risk. Thus, leukemia was about 1.3 times more likely after receipt of vitamin K, but this was not statistically significant.[3]

3. **Matched case–control study.**
(To illustrate the similarity between analysis of a matched case–control study and a case-crossover study, we will use the same example for both.) The research question is whether mobile telephone use increases the risk of car crashes among mobile telephone owners. A traditional matched case–control study might consider self-reported frequency of using a mobile telephone while driving as the risk factor. Then the cases would be people who had been in crashes and they could be compared with controls who had not been in crashes, matched by age, sex, and mobile telephone prefix to the cases. The cases and controls would then be asked whether they ever use a mobile telephone while driving. (To simplify, for this example, we dichotomize the exposure and consider people as either "users" or "nonusers" of mobile telephones while driving.) We then classify each case/control pair according to whether both are users, neither is a user, or the case was a user but not the control, or the control was a user but not the case. If we had 300 pairs, the results might look like this:

TABLE 8A.3

	CASES (WITH CRASH INJURIES)		
MATCHED CONTROLS	USER	NONUSER	TOTAL
User	110	40	150
Nonuser	90	60	150
Total	200	100	300

Table 8A.3 shows that there were 90 pairs where the case ever used a mobile phone while driving, but not the matched control, and 40 pairs where the matched control but not the case was a "user." Note that this 2 × 2 table is different from the 2 × 2 table from the unmatched vitamin K study in question 2, in which each cell in the table is the number of people in that cell. In the 2 × 2 table for a *matched* case–control study the number in each cell is the number of *pairs* of subjects in that cell; the total *N* in Table 8A.3 is therefore 600 (300 cases and 300 controls). The odds ratio for such a table is simply the ratio of the two types of discordant pairs; in the Table 8.A.3 the OR = 90/40 = 2.25. This implies that users of mobile phones had more than double the odds of being in a crash.

4. **Case-crossover study.** Now consider the case-crossover study of the same question. Data from the study by McEvoy et al. are shown below.

TABLE 8A.4

	CRASH TIME PERIOD		
SEVEN DAYS BEFORE CRASH	DRIVER USING PHONE	NOT USING	TOTAL
Driver using phone	5	6	11
Not using	27	288	315
Total	32	294	326

[3]The authors actually did a multivariate, matched analysis, as was appropriate for the matched design, but in this case the simple, unmatched odds ratio was almost the same as the one reported in the study.

For the case-crossover study, each cell in the table is a number of subjects, not a number of pairs, but *each cell represents two time periods* for that one subject: the time period just before the crash and a comparison time period 7 days before. Therefore, the 5 in the upper left cell means there were 5 drivers involved in crashes who were using a mobile phone just before they crashed, and also using a mobile phone during the comparison period 7 days before, while the 27 just below the 5 indicates that there were 27 drivers involved in crashes who were using a phone just before crashing, but *not* using a phone during the comparison period 7 days before. Similarly, there were 6 drivers involved in crashes who were not using their phone at the time of the crash, but were using them in the comparison time period 7 days before. The odds ratio is the ratio of the numbers of discordant time periods, in this example $27/6 = 4.5$, meaning that driving during time periods of mobile phone use is associated with 4.5-fold higher odds of a crash than driving during time periods when not using a mobile phone.

APPENDIX 8B
Why the Odds Ratio Can Be Used as an Estimate for Relative Risk in a Case–Control Study

The data in a case–control study represent two samples: The cases are drawn from a population of people who have the disease and the controls from a population of people who do not have the disease. The predictor variable (risk factor) is measured, and the results can be summarized in a 2 × 2 table like the following one:

	Cases	Controls
Risk factor present	*a*	*b*
Risk factor absent	*c*	*d*

If this 2 × 2 table represented data from a cohort study, then the incidence of the disease in those with the risk factor would be $a/(a + b)$ and the relative risk would be simply $[a/(a + b)]/[c/(c + d)]$. However, it is not appropriate to compute either incidence or relative risk in this way in a case–control study because the two samples are not drawn from the population in the same proportions. Usually, there are roughly equal numbers of cases and controls in the study samples but many fewer cases than controls in the population. Instead, relative risk in a case–control study can be approximated by the odds ratio, computed as the cross-product of the 2 × 2 table, ad/cb.

This extremely useful fact is difficult to grasp intuitively but easy to demonstrate algebraically. Consider the situation for the full population, represented by a', b', c', and d'.

	Disease	No Disease
Risk factor present	a'	b'
Risk factor absent	c'	d'

Here it is appropriate to calculate the risk of disease among people with the risk factor as $a'/(a' + b')$, the risk among those without the risk factor as $c'/(c' + d')$, and the relative risk as $[a'/(a' + b')]/[c'/(c' + d')]$. We have already discussed the fact that $a'/(a' + b')$ is not equal to $a/(a + b)$. However, if the disease is relatively uncommon in both those with and without the risk factor (as most are), then a' is much smaller than b', and c' is much smaller than d'. This means that $a'/(a' + b')$ is closely approximated by a'/b' and that $c'/(c' + d')$ is closely approximated by c'/d'. Therefore, the relative risk of the population can be approximated as follows:

$$\frac{a'/(a' + b')}{c'/(c' + d')} \approx \frac{a'/b'}{c'/d'}$$

The latter term is the odds ratio of the population (literally, the ratio of the odds of disease in those with the risk factor, a'/b', to the odds of disease in those without the risk factor, c'/d'). This can be rearranged as the cross-product:

$$\left(\frac{a'}{c'}\right)\left(\frac{d'}{c'}\right) = \left(\frac{a'}{c'}\right)\left(\frac{d'}{b'}\right)$$

However, a'/c' in the population equals a/c in the sample if the cases are representative of all cases in the population (i.e., have the same prevalence of the risk factor). Similarly, b'/d' equals b/d if the controls are representative.

Therefore, the population parameters in this last term can be replaced by the sample parameters, and we are left with the fact that the odds ratio observed in the sample, ad/bc, is a close approximation of the relative risk in the population, $[a'/(a' + b')]/[c'/(c' + d')]$, provided that the disease is rare.

REFERENCES

1. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971;284(15):878–881.
2. Beal SM, Finch CF. An overview of retrospective case–control studies investigating the relationship between prone sleeping position and SIDS. *J Paediatr Child Health* 1991;27(6):334–339.
3. Mitchell EA, Hutchison L, Stewart AW. The continuing decline in SIDS mortality. *Arch Dis Child* 2007;92(7):625–626.
4. Golding J, Greenwood R, Birmingham K, Mott M. Childhood cancer, intramuscular vitamin K, and pethidine given during labour. *BMJ* 1992;305(6849):341–346.
5. Golding J, Paterson M, Kinlen LJ. Factors associated with childhood cancer in a national cohort study. *Br J Cancer* 1990;62(2):304–308.
6. von Kries R, Gobel U, Hachmeister A, et al. Vitamin K and childhood cancer: a population based case–control study in Lower Saxony, Germany. *BMJ* 1996;313(7051):199–203.
7. Fear NT, Roman E, Ansell P, et al. Vitamin K and childhood cancer: a report from the United Kingdom Childhood Cancer Study. *Br J Cancer* 2003;89(7):1228–1231.
8. Roman E, Fear NT, Ansell P, et al. Vitamin K and childhood cancer: analysis of individual patient data from six case-control studies. *Br J Cancer* 2002;86(1):63–69.
9. Kochen M, McCurdy S. Circumcision and the risk of cancer of the penis. A life-table analysis. *Am J Dis Child* 1980;134(5):484–486.
10. O'Brien KL, Selanikio JD, Hecdivert C, et al. Epidemic of pediatric deaths from acute renal failure caused by diethylene glycol poisoning. Acute Renal Failure Investigation Team. *JAMA* 1998;279(15):1175–1180.
11. Fatal poisoning among young children from diethylene glycol-contaminated acetaminophen - Nigeria, 2008–2009. *MMWR Morb Mortal Wkly Rep* 2009;58(48):1345–1347.
12. Puumala SE, Spector LG, Robison LL, et al. Comparability and representativeness of control groups in a case–control study of infant leukemia: a report from the Children's Oncology Group. *Am J Epidemiol* 2009;170(3):379–387.
13. Voigt LF, Schwartz SM, Doody DR, et al. Feasibility of including cellular telephone numbers in random digit dialing for epidemiologic case–control studies. *Am J Epidemiol* 2011;173(1):118–126.
14. Hurwitz ES, Barrett MJ, Bregman D, et al. Public Health Service study of Reye's syndrome and medications. Report of the main study. *JAMA* 1987;257(14):1905–1911.
15. Cockburn M, Hamilton A, Mack T. Recall bias in self-reported melanoma risk factors. *Am J Epidemiol* 2001; 153(10):1021–1026.
16. Parr CL, Hjartaker A, Laake P, et al. Recall bias in melanoma risk factors and measurement error effects: a nested case-control study within the Norwegian Women and Cancer Study. *Am J Epidemiol* 2009;169(3):257–266.
17. Gefeller O. Invited commentary: Recall bias in melanoma—much ado about almost nothing? *Am J Epidemiol* 2009;169(3):267–270; discussion 71–72.
18. Cummings SR. Are patients with hip fractures more osteoporotic? Review of the evidence. *Am J Med* 1985;78(3):487–494.
19. Cauley JA, Lucas FL, Kuller LH, et al. Elevated serum estradiol and testosterone concentrations are associated with a high risk for breast cancer. Study of Osteoporotic Fractures Research Group. *Ann Intern Med* 1999; 130(4 Pt 1):270–277.
20. Azoulay L, Yin H, Filion KB, et al. The use of pioglitazone and the risk of bladder cancer in people with type 2 diabetes: nested case–control study. *BMJ* 2012;344:e3645.
21. Hassey A, Gerrett D, Wilson A. A survey of validity and utility of electronic patient records in a general practice. *BMJ* 2001;322(7299):1401–1405.
22. Essebag V, Platt RW, Abrahamowicz M, et al. Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. *BMC Med Res Methodol* 2005;5(1):5.
23. McEvoy SP, Stevenson MR, McCartt AT, et al. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ* 2005;331(7514):428.
24. Bhaskaran K, Hajat S, Armstrong B, et al. The effects of hourly differences in air pollution on the risk of myocardial infarction: case crossover analysis of the MINAP database. *BMJ* 2011;343:d5531.

25. Nuvolone D, Balzi D, Chini M, et al. Short-term association between ambient air pollution and risk of hospitalization for acute myocardial infarction: results of the cardiovascular risk and air pollution in Tuscany (RISCAT) study. *Am J Epidemiol* 2011;174(1):63–71.
26. Tramuto F, Cusimano R, Cerame G, et al. Urban air pollution and emergency room admissions for respiratory symptoms: a case-crossover study in Palermo, Italy. *Environ Health* 2011;10:31.
27. Scheers H, Mwalili SM, Faes C, et al. Does air pollution trigger infant mortality in Western Europe? A case-crossover study. *Environ Health Perspect* 2011;119(7):1017–1022.
28. Reijneveld SA, Brugman E, Hirasing RA. Infantile colic: maternal smoking as potential risk factor. *Arch Dis Child* 2000;83(4):302–303.