3

Basic Principles of Clinical Epidemiology Relevant to Pharmacoepidemiologic Studies

Brian L. Strom

Rutgers Biomedical and Health Sciences, Newark, NJ, USA

Pharmacoepidemiology applies the methods of epidemiology to the content area of clinical pharmacology. Chapter 2 reviewed the basic principles of clinical pharmacology. Therefore, in order to understand the approaches and methodologic issues specific to the field of pharmacoepidemiology, the basic principles of epidemiology must be understood as well. To that end, this chapter will begin with an overview of the scientific method in general. This will be followed by a discussion of the different types of errors one can make in designing a study. Next, the chapter will review the criteria for the causal nature of an association, which are how one can decide whether an association demonstrated in a particular study is, in fact, a causal association. Finally, the specific study designs available for epidemiologic studies, or in fact for any clinical studies, will be reviewed. The next chapter discusses a specific methodologic issue which needs to be addressed in any study, but which is of particular importance for pharmacoepidemiologic studies: the issue of sample size. These two chapters are intended to be an introduction to the field of epidemiology for the neophyte. More information on these principles can be obtained from any textbook on epidemiology or clinical epidemiology [1–24].

Overview of the Scientific Method

The scientific method to investigate a research question involves a three-stage process (see Figure 3.1). In the first stage, one selects a group of subjects for study. These subjects may be patients or animals or biologic cells, and are the sources for the data sought by the study to answer a question of interest. Second, one uses the information obtained in this sample of study subjects to generalize and draw a conclusion about a population in general. This conclusion is referred to as an association. Third, one generalizes again, drawing a conclusion about a scientific theory or causation. Each will be discussed in turn.

Any given study is performed on a selection of individuals, who represent the *study subjects*. These study subjects should theoretically represent a random sample of some defined population. For example, one might perform a randomized clinical trial of the efficacy of enalapril in lowering blood pressure, randomly allocating a total of 40 middle-aged hypertensive men to receive either enalapril or placebo and observing their blood pressure six weeks later. One might expect to see the blood

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

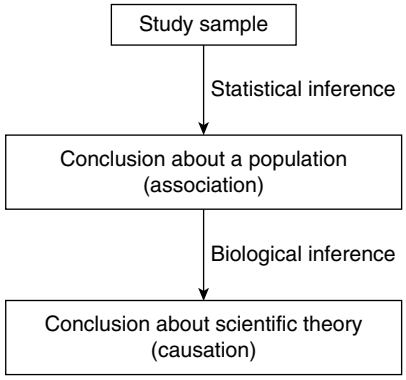

Figure 3.1 Overview of the scientific method.

pressure of the 20 men treated with the active drug decrease more than the blood pressure of the 20 men treated with a placebo. In this example, the 40 study subjects would represent the study sample, theoretically a random sample of middle-aged hypertensive men. In reality, the study sample is almost never a true random sample of the underlying target population, because it is logistically impossible to identify every individual who belongs in the target population and then randomly choose from among them. However, the study sample is usually treated as if it were a random sample of the target population.

At this point, one would be tempted to make a generalization that enalapril lowers blood pressure in middle-aged hypertensive men. However, one must explore whether this observation could have occurred simply by chance; that is, due to random variation. If the observed outcome in the study was simply a chance occurrence, then the same observation might not have been seen if one had chosen a different sample of 40 study subjects. Perhaps more importantly, it might not exist if one were able to study the entire theoretical population of all middle-aged hypertensive men. In order to evaluate this possibility, one can perform a statistical test, which allows an investigator to quantitate the probability that the observed outcome in this study (i.e., the difference seen between the two study groups) could have happened simply by chance. There are explicit rules and procedures for how one should properly make this determination: the science of statistics. If the results of any study under consideration demonstrate a "statistically significant difference" (i.e., ruling out the probability of a chance occurrence), then one is said to have an *association*. The process of assessing whether random variation could have led to a study's findings is referred to as *statistical inference*, and represents the major role for statistical testing in the scientific method.

If there is no statistically significant difference, then the process in Figure 3.1 stops. If there is an association, then one is tempted to generalize the results of the study even further, to state that enalapril is an antihypertensive drug in general. This is referred to as *scientific or biologic inference*, and the result is a conclusion about *causation*, that the drug really does lower blood pressure in a population of treated patients. To draw this type of conclusion, however, requires one to generalize to populations other than that included in the study, including types of people who were not represented in the study sample, such as women, children, and the elderly. Although it may be apparent in this example that this is in fact appropriate, that may well not always be the case. Unlike statistical inference, there are no precise quantitative rules for biologic inference. Rather, one needs to examine the data at hand in light of all other relevant data in the rest of the scientific literature, and make a subjective judgment. To assist in making that judgment, however, one can use the criteria for the causal nature of an association described later in the chapter. First, however, we will place causal associations into proper perspective by describing the different types of errors that can be made in performing a study and the different types of associations in which each results.

Types of Errors That One Can Make in Performing a Study

There are four basic types of associations that can be observed in a study (Table 3.1). The basic purpose of research is to differentiate among them.

First, of course, one could have no association.

Second, one could have an *artifactual association*; that is, a spurious or false association. This can occur by either of two mechanisms: chance or bias. Chance is unsystematic, or random, variation. The purpose of statistical testing in science is to evaluate this, estimating the probability that the result observed in a study could have happened purely by chance.

The other possible mechanism for creating an artifactual association is bias. Epidemiologists' use of the term bias is different from that of the lay public. To an epidemiologist, *bias* is systematic variation, a consistent manner in which two study groups are treated or evaluated differently. This consistent difference can create an apparent association where one actually does not exist. Of course, it also can mask a true association.

There are many different types of potential biases [25]. For example, consider an interview study in which the research assistant is aware of the investigator's hypothesis. Attempting to please the boss, the research assistant might probe more carefully during interviews with one study group than during interviews with the other. This difference in how carefully the interviewer probes could create an apparent but false association, which is referred to as interviewer bias. Another example would be a study of drug-induced birth defects that compares children with birth defects to children without birth defects. A mother of a child with birth defect, when interviewed about any drugs she took during her pregnancy, may be likely to remember drug ingestion during pregnancy with greater accuracy than a mother of a healthy child, because of the unfortunate experience she has undergone. The improved recall in the mothers of the children with birth defects may result in false apparent associations between drug exposure and birth defects. This systematic difference in recall is referred to as recall bias [26].

Table 3.1 Types of associations between factors under study.

1) None (independent)
2) Artifactual (spurious or false)
   a) Chance (unsystematic variation)
   b) Bias (systematic variation)
3) Indirect (confounded)
4) Causal (direct or true)

Note that biases, once present, cannot be corrected. They represent errors in the study design that can result in incorrect results in the study. It is important to note that a *statistically significant result is no protection against a bias*; one can have a very precise measurement of an incorrect answer! The only protection against biases is proper study design. (See Chapter 43 for more discussion about biases in pharmacoepidemiologic studies.)

Third, one can have an indirect, or confounded, association. A *confounding variable*, or *confounder*, is a variable, other than the risk factor and other than the outcome under study, which is related independently to both the risk factor and the outcome and which may create an apparent association or mask a real one. For example, a study of risk factors for lung cancer could find a very strong association between having yellow fingertips and developing lung cancer. This is obviously not a causal association, but an indirect association, confounded by cigarette smoking. Specifically, cigarette smoking causes both yellow fingertips and lung cancer. Although this example is transparent, most examples of confounding are not. In designing a study, one must consider every variable that can be associated with the risk factor under study or the outcome variable under study, in order to

Table 3.2 Approaches to controlling confounding.

1) Random allocation
2) Subject selection
a) Exclusion
b) Matching
3) Data analysis
a) Stratification
b) Mathematical modeling

plan to deal with it as a potential confounding variable. Preferably, one will be able to specifically control for the variable, using one of the techniques listed in Table 3.2. (See Chapters 33 and 43 for more discussion about confounding in pharmacoepidemiologic studies.)

Fourth, and finally, there are true, causal associations.

Thus, there are three possible types of errors that can be produced in a study: random error, bias, and confounding. The probability of random error can be quantitated using statistics. Bias needs to be prevented by designing the study properly. Confounding can be controlled either in the design of the study or in its analysis. If all three types of errors can be excluded, then one is left with a true, causal association.

Criteria for the Causal Nature of an Association

The "criteria for the causal nature of an association" were first put forth by Sir Austin Bradford Hill [27], but have been described in various forms since, each with some modification. Probably the best known description of them was in the first Surgeon General's Report on Smoking and Health [28], published in 1964. These criteria are presented in Table 3.3, in no particular order. No one of them is absolutely necessary for an association to be a causal association. Analogously, no one of them is sufficient for an association to be considered a causal association. Essentially, the more criteria that are present, the more likely it is that an association is a causal association. The fewer criteria that are met, the less likely it is that an association is a causal association. Each will be discussed in turn.

Table 3.3 Criteria for the causal nature of an association.

1) Coherence with existing information (biologic plausibility)
2) Consistency of the association
3) Time sequence
4) Specificity of the association
5) Strength of the association
a) Quantitative strength
b) Dose–response relationship
c) Study design

The first criterion listed in Table 3.3 is *coherence with existing information* or *biological plausibility*. This refers to whether the association makes sense, in light of other types of information available in the literature. These other types of information could include data from other human studies, data from studies of other related questions, data from animal studies, or data from *in vitro* studies, as well as scientific or pathophysiologic theory. To use the example provided earlier, it clearly was not biologically plausible that yellow fingertips could cause lung cancer, and this provided the clue that confounding was present. Using the example of the association between cigarettes and lung cancer, cigarette smoke is a known carcinogen, based on animal data. In humans, it is known to cause cancers of the head and neck, the pancreas, and the bladder. Cigarette smoke also goes down into the lungs, directly exposing the tissues in question. Thus, it certainly is biologically plausible that cigarettes could *cause* lung cancer [29]. It is much more reassuring if an association found in a particular study makes sense, based on previously available information, and this leads one to be more comfortable that it

might be a causal association. Clearly, however, one could not require that this criterion always be met, or one would never have a major breakthrough in science.

The second criterion listed in Table 3.3 is the *consistency of the association*. A hallmark of science is reproducibility: if a finding is real, one should be able to reproduce it in a different setting. This could include different geographic settings, different study designs, different populations, and so on. For example, in the case of cigarettes and lung cancer, the association has now been reproduced in many different studies, in different geographic locations, using different study designs [30]. The need for reproducibility is such that one should never believe a finding reported only once; there may have been an error committed in the study, which is not apparent to either the investigator or the reader.

The third criterion listed is the *time sequence*: a cause must precede an effect. Although this may seem obvious, there are study designs from which this cannot be determined. For example, if one were to perform a survey in a classroom of 200 medical students, asking each if they were currently taking diazepam and also whether they were anxious, one would find a strong association between the use of diazepam and anxiety, but this does not mean that diazepam causes anxiety! Although this is obvious, as it is not a biologically plausible interpretation, one cannot differentiate from this type of cross-sectional study which variable came first and which came second. In the example of cigarettes and lung cancer, obviously the cigarette smoking usually precedes the lung cancer, as a patient would not survive long enough to smoke much if the opposite were the case.

The fourth criterion listed in Table 3.3 is *specificity*. This refers to the question of whether the cause ever occurs without the presumed effect, and whether the effect ever occurs without the presumed cause. This criterion is almost never met in biology, with the occasional exception of infectious diseases. Measles never occurs without the measles virus, but even in this example, not everyone who becomes infected with the measles virus develops clinical measles. Certainly, not everyone who smokes develops lung cancer, and not everyone who develops lung cancer was a smoker. This is one of the major points the tobacco industry stresses when it attempts to make the claim that cigarette smoking has not been proven to cause lung cancer. Some authors even omit this as a criterion, as it is so rarely met. When it is met, however, it provides extremely strong support for a conclusion that an association is causal.

The fifth criterion listed in Table 3.3 is the *strength of the association*. This includes three concepts: its quantitative strength, dose–response, and the study design. Each will be discussed in turn.

The *quantitative strength* of an association refers to the effect size. To evaluate this, one asks whether the magnitude of the observed difference between the two study groups is large. A quantitatively large association can only be created by a causal association or a large error, which should be apparent in evaluating the methods of a study. A quantitatively small association may still be causal, but it could be created by a subtle error, which would not be apparent in evaluating the study. Conventionally, epidemiologists consider an association with a relative risk of less than 2.0 a weak association. Certainly, the association between cigarette smoking and lung cancer is a strong association: studies show relative risks ranging between 10.0 and 30.0 [30].

A *dose–response relationship* is an extremely important and commonly used concept in clinical pharmacology and is used similarly in epidemiology. It exists when an increase in the intensity of an exposure results in an increased risk of the disease under study. Equivalent to this is a *duration–response relationship*, which exists when a longer exposure causes an increased risk of the disease. The presence of either relationship strongly implies that an

association is, in fact, a causal association. Certainly in the example of cigarette smoking and lung cancer, it has been shown repeatedly that an increase in either the number of cigarettes smoked each day or in the number of years of smoking increases the risk of developing lung cancer [30].

Finally, *study design* refers to two concepts: whether the study was well designed, and which study design was used in the studies in question. The former refers to whether the study was subject to one of the three errors described earlier in this chapter, namely random error, bias, and confounding. Table 3.4 presents the study designs typically used for epidemiologic studies, or in fact for any clinical studies. They are organized in a hierarchical fashion. As one advances from the designs at the bottom of the

Table 3.4 Advantages and disadvantages of epidemiologic study designs.

Study design	Advantages	Disadvantages
Randomized clinical trial (experimental study)	Most convincing design Only design which controls for unknown or unmeasurable confounders	Most expensive Artificial Logistically most difficult Ethical objections
Cohort study	Can study multiple outcomes Can study uncommon exposures Selection bias less likely Unbiased exposure data Incidence data available	Possibly biased outcome data More expensive If done prospectively, may take years to complete
Case–control study	Can study multiple exposures Can study uncommon diseases Logistically easier and faster Less expensive	Control selection problematic Possibly biased exposure data
Analyses of secular trends	Can provide rapid answers	No control of confounding
Case series	Easy quantitation of incidence	No control group, so cannot be used for hypothesis testing
Case reports	Cheap and easy method for generating hypotheses	Cannot be used for hypothesis testing

table to those at the top, studies get progressively harder to perform, but are progressively more convincing. In other words, associations shown by studies using designs at the top of the list are more likely to be causal associations than associations shown by studies using designs at the bottom of the list. The association between cigarette smoking and lung cancer has been reproduced in multiple well-designed studies, using analyses of secular trends, case–control studies, and cohort studies. However, it has not been shown using a randomized clinical trial, which is the "Cadillac" of study designs, as will be discussed later in the chapter. This is the other major defense the tobacco industry employs. Of course, it would not be ethical or logistically feasible to randomly allocate individuals to smoke or not to smoke and expect to follow them for 20 years to observe the outcome in each group.

The issue of causation is discussed more in Chapter 10 as it relates to the process of spontaneous reporting of adverse drug reactions, and in Chapter 29 as it relates to determining causation in case reports.

Epidemiologic Study Designs

In order to clarify the concept of study design further, each of the designs in Table 3.4 will be discussed in turn, starting at the bottom of the list and working upward.

Case Reports

Case reports are simply reports of events observed in single patients. As used in pharmacoepidemiology, a case report describes a single patient who was exposed to a drug and experiences a particular, usually adverse, outcome. For example, one might see a published case report about a young woman who was taking oral contraceptives and who suffered a pulmonary embolism.

Case reports are useful for raising hypotheses about drug effects, to be tested with more rigorous study designs. However, in a case report one cannot know if the patient reported is either typical of those with the exposure or typical of those with the disease. Certainly, one cannot usually determine whether the adverse outcome was due to the drug exposure or would have happened anyway. As such, it is very rare that a case report can be used to make a statement about causation. One exception to this would be when the outcome is so rare and so characteristic of the exposure that one knows that it was likely to be due to the exposure, even if the history of exposure were unclear. An example of this is clear cell vaginal adenocarcinoma occurring in young women exposed *in utero* to diethylstilbestrol [31]. Another exception would be when the disease course is very predictable and the treatment causes a clearly apparent change in this disease course. An example would be the ability of penicillin to cure streptococcal endocarditis, a disease that is nearly uniformly fatal in the absence of treatment. Case reports can be particularly useful to document causation when the treatment causes a change in disease course which is reversible, such that the patient returns to their untreated state when the exposure is withdrawn, can be treated again, and when the change returns upon repeat treatment. Consider a patient who is suffering from an overdose of methadone, a long-acting narcotic, and is comatose. If this patient is then treated with naloxone, a narcotic antagonist, and immediately awakens, this would be very suggestive that the drug indeed is efficacious as a narcotic antagonist. As the naloxone wears off the patient would become comatose again, and then if they were given another dose of naloxone they would awaken again. This, especially if repeated a few times, would represent strong evidence that the drug is indeed effective as a narcotic antagonist. This type of challenge–rechallenge situation is relatively uncommon, however, as physicians generally will avoid

exposing a patient to a drug if the patient experienced an adverse reaction to it in the past. This issue is discussed in more detail in Chapters 10 and 29.

Case Series

Case series are collections of patients, all of whom have a single exposure, whose clinical outcomes are then evaluated and described. Often they are from a single hospital or medical practice. Alternatively, case series can be collections of patients with a single outcome, looking at their antecedent exposures. For example, one might observe 100 consecutive women under the age of 50 who suffer from a pulmonary embolism, and note that 30 of them had been taking oral contraceptives.

After drug marketing, case series are most useful for two related purposes. First, they can be useful for quantifying the incidence of an adverse reaction. Second, they can be useful for being certain that any particular adverse effect of concern does not occur in a population which is larger than that studied prior to drug marketing. The so-called Phase IV postmarketing surveillance study of prazosin was conducted for the former reason, to quantitate the incidence of first-dose syncope from prazosin [32]. The Phase IV postmarketing surveillance study of cimetidine [33] was conducted for the latter reason. Metiamide was an H-2 blocker, which was withdrawn after marketing outside the US because it caused agranulocytosis. Since cimetidine is chemically related to metiamide, there was a concern that cimetidine too might cause agranulocytosis [32]. In both examples, the manufacturer asked its sales representatives to recruit physicians to participate in the study. Each participating physician then enrolled the next series of patients for whom the drug was prescribed.

In this type of study, one can be more certain that the patients are probably typical of those with the exposure or with the disease, depending on the focus of the study. However, in the absence of a control group, one cannot be certain which features in the description of the patients are unique to the exposure or outcome. As an example, one might have a case series from a particular hospital of 100 individuals with a certain disease, and note that all were men over the age of 60. This might lead one to conclude that this disease seems to be associated with being a man over the age of 60. However, it would be clear that this would be an incorrect conclusion once one noted that the hospital this case series was drawn from was a Veterans Administration hospital, where most patients are men over the age of 60. In the previous example of pulmonary embolism and oral contraceptives, 30% of the women with pulmonary embolism had been using oral contraceptives. However, this information is not sufficient to determine whether this is higher, the same as, or even lower than would have been expected. For this reason, case series are also not very useful in determining causation, but provide clinical descriptions of a disease or of patients who receive an exposure.

Analyses of Secular Trends

Analyses of secular trends, also called "ecologic studies," examine trends in an exposure that is a presumed cause and trends in a disease that is a presumed effect and test whether the trends coincide. These trends can be examined over time or across geographic boundaries. In other words, one could analyze data from a single region and examine how the trend changes over time, or one could analyze data from a single time period and compare how the data differ from region to region or country to country. Vital statistics are often used for these studies. As an example, one might look at sales data for oral contraceptives and compare them to death rates from venous thromboembolism, using recorded vital statistics. When such a study was actually performed, mortality rates from venous

thromboembolism were seen to increase in parallel with increasing oral contraceptive sales, but only in women of reproductive age, not in older women or in men of any age [34].

Analyses of secular trends are useful for rapidly providing evidence for or against a hypothesis. However, these studies lack data on individuals; they utilize only aggregated group data (e.g., annual sales data in a given geographic region in relation to annual cause-specific mortality in the same region). As such, they are unable to control for confounding variables. Thus, among exposures whose trends coincide with that of the disease, analyses of secular trends are unable to differentiate which factor is likely to be the true cause. For example, lung cancer mortality rates in the US have been increasing in women, such that lung cancer is now the leading cause of cancer mortality in women [35]. This is certainly consistent with the increasing rates of cigarette smoking observed in women until the mid-1960s [36], and so appears to be supportive of the association between cigarette smoking and lung cancer. However, it would also be consistent with an association between certain occupational exposures and lung cancer, as more women in the US are now working outside the home.

Case–Control Studies

Case–control studies compare cases with a disease to controls without the disease, looking for differences in antecedent exposures. As an example, one could select cases of young women with venous thromboembolism and compare them to controls without venous thromboembolism, looking for differences in antecedent oral contraceptive use. Several such studies have been performed, generally demonstrating a strong association between the use of oral contraceptives and venous thromboembolism [37].

Case–control studies can be particularly useful when one wants to study multiple possible causes of a single disease, as one can use the same cases and controls to examine any number of exposures as potential risk factors. This design is also particularly useful when one is studying a relatively rare disease, as it guarantees a sufficient number of cases with the disease. Using case–control studies, one can study rare diseases with markedly smaller sample sizes than those needed for cohort studies (see Chapter 4). For example, the classic study of diethylstilbestrol and clear cell vaginal adenocarcinoma required only 8 cases and 40 controls [31], rather than the many thousands of exposed subjects that would have been required for a cohort study of this question.

Case–control studies generally obtain their information on exposures retrospectively; that is, by recreating events that happened in the past. Information on past exposure to potential risk factors is generally obtained by abstracting medical records or by administering questionnaires or interviews. As such, case–control studies are subject to limitations in the validity of retrospectively collected exposure information. In addition, the proper selection of controls can be a challenging task and appropriate control selection can lead to a selection bias, which may lead to incorrect conclusions. Nevertheless, when case–control studies are done well, subsequent well-done cohort studies or randomized clinical trials, if any, will generally confirm their results. As such, the case–control design is a very useful approach for pharmacoepidemiologic studies.

Cohort Studies

Cohort studies identify subsets of a defined population and follow them over time, looking for differences in their outcome. Cohort studies generally are used to compare exposed patients to unexposed patients, although they can also be used to compare one exposure to another. For example, one could compare women of reproductive age who use oral contraceptives to users of other contraceptive methods, looking

for the differences in the frequency of venous thromboembolism. When such studies were performed, they in fact confirmed the relationship between oral contraceptives and thromboembolism, which had been noted using analyses of secular trends and case–control studies [38,39]. Cohort studies can be performed either prospectively, that is simultaneous with the events under study, or retrospectively, that is after the outcomes under study had already occurred, by recreating those past events using medical records, questionnaires, or interviews.

The major difference between cohort and case–control studies is the basis upon which patients are recruited into the study (see Figure 3.2). Patients are recruited into case–control studies based on the presence or absence of a disease, and their antecedent exposures are then studied. Patients are recruited into cohort studies based on the presence or absence of an exposure, and their subsequent disease course is then studied.

Cohort studies have the major advantage of being free of the major problem that plagues case–control studies: the difficult process of selecting an undiseased control group. In addition, prospective cohort studies are free of the problem of the questionable validity of retrospectively collected data. For these reasons, an association demonstrated by a cohort study is more likely to be a causal association than one demonstrated by a case–control study. Furthermore, cohort studies are particularly useful when one is studying multiple possible outcomes from a single exposure, especially a relatively uncommon exposure. Thus, they are especially useful in postmarketing drug surveillance studies, which are looking at any possible effect of a newly marketed drug. However, cohort studies can require extremely large sample sizes to study relatively uncommon outcomes (see Chapter 4). In addition, prospective cohort studies can require a prolonged time period to study delayed drug effects.

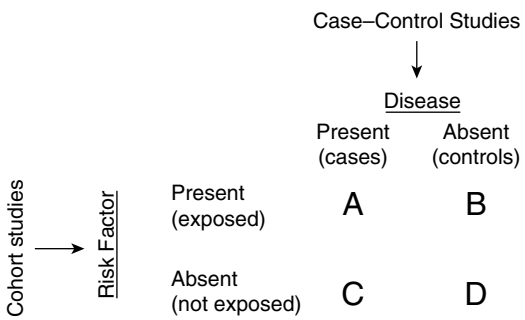

Figure 3.2 Cohort and case–control studies provide similar information, but approach data collection from opposite directions. *Source:* Strom BL. Medical databases in post-marketing drug surveillance. *Trends Pharmacol Sci* 1986; **7**: 377–80.

Analysis of Case–Control and Cohort Studies

As can be seen in Figure 3.2, both case–control and cohort studies are intended to provide the same basic information; the difference is how this information is collected. The key statistic reported from these studies is the *relative risk*, the ratio of the incidence rate of an outcome in the exposed group to the incidence rate of the outcome in the unexposed group. A relative risk of greater than 1.0 means that exposed subjects have a *greater* risk of the disease under study than unexposed subjects, or that the exposure appears to cause the disease. A relative risk of less than 1.0 means that exposed subjects have a *lower* risk of the disease than unexposed subjects, or that the exposure seems to protect against the disease. A relative risk of 1.0 means that exposed subjects and unexposed subjects have the same risk of developing the disease, or that the exposure and the disease appear unrelated.

One can calculate a relative risk directly from the results of a cohort study. However, in a case–control study one cannot determine the size of either the exposed population or the unexposed population from which the diseased cases and undiseased controls were drawn. The results of a case–control study do not provide

information on the incidence rates of the disease in exposed and unexposed individuals. Therefore, relative risks cannot be calculated directly from a case–control study. Instead, in reporting the results of a case–control study one generally reports the *odds ratio*, which is a close estimate of the relative risk when the disease under study is relatively rare. Since case–control studies are generally used to study rare diseases, there generally is very close agreement between the odds ratio and the relative risk, and the results from case–control studies are often loosely referred to as relative risks, although they are in fact odds ratios.

Both relative risks and odds ratios can be reported with *P values*. These P values allow one to determine if the relative risk is statistically significantly different from 1.0; that is, whether the differences between the two study groups are likely to be due to random variation or are likely to represent real associations.

Alternatively, and probably preferably, relative risks and odds ratios can be reported with *confidence intervals*, which are an indication of the range of relative risks within which the true relative risk for the entire theoretical population is most likely to lie. As an approximation, a 95% confidence interval around a relative risk means that we can be 95% confident that the true relative risk lies in the range between the lower and upper limits of this interval. If a 95% confidence interval around a relative risk excludes 1.0, then the finding is statistically significant with a P value of less than 0.05. A confidence interval provides much more information than a P value, however. As an example, a study that yields a relative risk (95% confidence interval) of 1.0 (0.9–1.1) is clearly showing that an association is very unlikely. A study that yields a relative risk (95% confidence interval) of 1.0 (0.1–100) provides little evidence for or against an association. Yet, both could be reported as a relative risk of 1.0 and a P value greater than 0.05. As another example, a study that yields a relative risk (95% confidence interval) of 10.0 (9.8–10.2) precisely quantifies a 10-fold increase in risk that is also statistically significant. A study that yields a relative risk (95% confidence interval) of 10.0 (1.1–100) says little, other than that an increased risk is likely. Yet, both could be reported as a relative risk of 10.0 ($P<0.05$). As a final example, a study yielding a relative risk (95% confidence interval) of 3.0 (0.98–5.0) is strongly suggestive of an association, whereas a study reporting a relative risk (95% confidence interval) of 3.0 (0.1–30) would not be. Yet, both could be reported as a relative risk of 3.0 ($P>0.05$).

Finally, another statistic that one can calculate from a cohort study is the excess risk, also called the risk difference or, sometimes, the attributable risk. Whereas the relative risk is the ratio of the incidence rates in the exposed group versus the unexposed groups, the excess risk is the arithmetic difference between the incidence rates. The relative risk is more important in considering questions of causation. The excess risk is more important in considering the public health impact of an association, as it represents the increased rate of disease due to the exposure. For example, oral contraceptives are strongly associated with the development of myocardial infarction in young women [37]. However, the risk of myocardial infarction in nonsmoking women in their 20s is so low, that even a fivefold increase in that risk would still not be of public health importance. In contrast, women in their 40s are at higher risk, especially if they are cigarette smokers as well. Thus, oral contraceptives should not be as readily used in these women [37].

As with relative risks, excess risks cannot be calculated from case–control studies, as incidence rates are not available. As with the other statistics, P values can be calculated to determine whether the differences between the two study groups could have occurred just by chance. Confidence intervals can be calculated around excess risks as well, and would be interpreted analogously.

Randomized Clinical Trials

Finally, *experimental studies* are studies in which the investigator controls the therapy that is to be received by each participant. Generally, an investigator uses that control to randomly allocate patients between or among the study groups, performing a *randomized clinical trial*. For example, one could theoretically randomly allocate sexually active women to use either oral contraceptives or no contraceptives, examining whether they differ in their incidence of subsequent venous thromboembolism. The major strength of this approach is random assignment, which is the only way to make it likely that the study groups are comparable in potential confounding variables that are either unknown or unmeasurable. For this reason, associations demonstrated in randomized clinical trials are more likely to be causal associations than those demonstrated using one of the other study designs reviewed here.

However, even randomized clinical trials are not without their problems. The randomized clinical trial just outlined, allocating women to receive contraceptives or no contraceptives, demonstrates the major potential problems inherent in the use of this study design. It would obviously be impossible to perform, ethically and logistically. In addition, randomized clinical trials are expensive and artificial. Inasmuch as they have already been performed prior to marketing to demonstrate each drug's efficacy, they tend to be unnecessary after marketing. They are likely to be used in pharmacoepidemiologic studies mainly for supplementary studies of drug efficacy [40]. However, they remain the "gold standard" by which the other designs must be judged. Indeed, with the publication of the results from the Women's Health Initiative indicating that combination hormone replacement therapy causes an increased risk of myocardial infarction rather than a decreased risk [41–44], there has been increased concern about reliance solely on nonexperimental methods to study drug safety after marketing [45–47], and we are seeing the use of massive randomized clinical trials as part of postmarketing surveillance (see Chapter 32).

Discussion

Thus, a series of different study designs are available (Table 3.4), each with their respective advantages and disadvantages. Case reports, case series, analyses of secular trends, case–control studies, and cohort studies have been referred to collectively as *observational study designs* or *nonexperimental study designs*, in order to differentiate them from experimental studies. In nonexperimental study designs the investigator does not control the therapy, but simply observes and evaluates the results of ongoing medical care. Case reports, case series, and analyses of secular trends have also been referred to as *descriptive studies*. Case–control studies, cohort studies, and randomized clinical trials all have control groups, and have been referred to as *analytic studies*. The analytic study designs can be classified in two major ways: by how subjects are selected into the study and by how data are collected for the study (see Table 3.5). From the perspective of how subjects are recruited into the study, case–control studies can be contrasted with cohort studies. Specifically, case–control studies select subjects into the study based on the presence or absence of a disease, while cohort studies select subjects into the study based on the presence or absence of an exposure. From this perspective, randomized clinical trials can be viewed as a subset of cohort studies, a type of cohort study in which the investigator controls the allocation of treatment, rather than simply observing ongoing medical care. From the perspective of timing, data can be collected *prospectively*, that is simultaneously with the events under study, or *retrospectively*, that is after the events under

study had already developed. In the latter situation, one recreates events that happened in the past using medical records, questionnaires, or interviews. Data can also be collected using *cross-sectional studies*, studies that have no time sense, as they examine only one point in time. In principle, either cohort or case–control studies can be performed using any of these time frames, although prospective case–control studies are unusual. Randomized clinical trials must be prospective, as this is the only way an investigator can control the therapy received.

The terms presented in this chapter, which are those that will be used throughout the book, are probably those used by a majority of epidemiologists. Unfortunately, however, other terms have been used for most of these study designs as well. Table 3.5 presents several of the synonyms that have been used in the medical literature. The same term is sometimes used by different authors to describe different concepts. For example, in this book we are reserving the use of the terms "retrospective study" and "prospective study" to refer to a time sense. As is apparent from Table 3.5, however, in the past some authors have used the term "retrospective study" to refer to a case–control study and "prospective study" to refer to a cohort study, confusing the two concepts inherent in the classification schemes presented in the table. Other authors use the term "retrospective study" to refer to any nonexperimental study, while others appear to use it to refer to any study they do not like, as a term of derision! Unfortunately, when reading a scientific paper, there is no way of determining which usage the author intended. More important than the terminology, however, are the concepts underlying the terms. Once they understand these concepts, readers can choose to use whatever terminology they are comfortable with.

Table 3.5 Epidemiologic study designs.

A) Classified by how subjects are recruited into the study
1) Case–control (case-history, case-referent, retrospective, trohoc) studies
2) Cohort (follow-up, prospective) studies
3) Experimental studies (clinical trials, intervention study)
B) Classified by how data are collected for the study
1) Retrospective (historical, nonconcurrent, retrolective) studies
2) Prospective (prolective) studies
3) Cross-sectional studies

Conclusion

From the material presented in this chapter, it is hopefully now apparent that each study design has an appropriate role in scientific progress. In general, science proceeds from the bottom of Table 3.4 upward, from case reports and case series that are useful for suggesting an association to analyses of trends and case–control studies that are useful for exploring these associations. Finally, if a study question warrants the investment and can tolerate the delay until results become available, then cohort studies and randomized clinical trials can be undertaken to assess these associations more definitively.

For example, regarding the question of whether oral contraceptives cause venous thromboembolism, an association was first suggested by case reports and case series, then was explored in more detail by analyses of trends and a series of case–control studies [37]. Later, because of the importance of oral contraceptives, the number of women using them, and the fact that users were predominantly healthy women, the investment was made in two long-term, large-scale cohort studies [38,39]. This question might even be worth the investment of a randomized clinical trial, except it would not be feasible or ethical. In contrast, when thalidomide was marketed, it was

not a major breakthrough; other hypnotics were already available. Case reports of phocomelia in exposed patients were followed by case–control studies [48] and analyses of secular trends [49]. Inasmuch as the adverse effect was so terrible and the drug was not of unique importance, the drug was then withdrawn, without the delay that would have been necessary if cohort studies and/or randomized clinical trials had been awaited. Ultimately, a retrospective cohort study was performed, comparing those exposed during the critical time period to those exposed at other times [50].

In general, however, clinical, regulatory, commercial, and legal decisions need to be made based on the best evidence available at the time of the decision. To quote Sir Austin Bradford Hill [27]:

> All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time. Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8:30 next day.

References

1 Schneider D, Lilienfeld DE. *Lilienfeld's Foundations of Epidemiology*, 4th edn. New York: Oxford University Press, 2014.
2 MacMahon B. *Epidemiology: Principles and Methods*, 2nd edn. Hagerstown, MD: Lippincott-Raven, 1997.
3 Friedman G. *Primer of Epidemiology*, 5th edn. New York: McGraw-Hill, 2003.
4 Mausner JS, Kramer S. *Epidemiology: An Introductory Text*, 2nd edn. Philadelphia, PA: Saunders, 1985.
5 Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*, 2nd edn. Boston, MA: Little Brown, 1991.
6 Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology, 3rd edn*. Philadelphia, PA: Lippincott Williams & Wilkins, 2012.
7 Weiss N. *Clinical Epidemiology: The Study of the Outcome of Illness*, 3rd edn. New York: Oxford University Press, 2006.
8 Kelsey JL, Whittemore AS, Evans AS. *Methods in Observational Epidemiology*, 2nd edn. New York: Oxford University Press, 1996.
9 Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston, MA: Little Brown, 1987.
10 Fletcher RH, Fletcher SW, Fletcher GS. *Clinical Epidemiology: The Essentials, 5th edn*. Lippincott Williams & Wilkins, 2012.
11 Gordis L. *Epidemiology*, 5th edn. Philadelphia, PA: Saunders, 2013.
12 Rothman KJ. *Epidemiology: An Introduction*, 2nd edn. New York: Oxford University Press, 2012.
13 Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*, 3rd edn. Sudbury, MA: Jones and Bartlett, 2014.
14 Katz DL, Wild D, Elmore JG, Lucan SC. *Jekel's Epidemiology, Biostatistics, and Preventive Medicine*, 4th edn. Philadelphia, PA: Elsevier, 2014.
15 Aschengrau A, Seage GR. *Essentials of Epidemiology in Public Health*, 3rd edn. Sudbury, MA: Jones and Bartlett, 2014.
16 Friis RH, Sellers TA. *Epidemiology for Public Health Practice*, 5th edn. Sudbury, MA: Jones & Bartlett, 2014.

17 Hulley SB, Cummings SR, Browner WS, Grady D, Newman TB. *Designing Clinical Research: An Epidemiologic Approach*, 4th edn. Baltimore, MD: Lippincott Williams & Wilkins, 2013.

18 Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring JR. *Medical Epidemiology*, 4th edn. New York: McGraw-Hill, 2004.

19 Wassertheil-Smoller S, Smoller J. *Biostatistics and Epidemiology: A Primer for Health Professionals*, 4th edn. New York: Springer, 2015.

20 Katz DL. *Clinical Epidemiology and Evidence-Based Medicine: Fundamental Principles of Clinical Reasoning and Research*. Thousand Oaks, CA: Sage, 2001.

21 Bhopal RS. *Concepts of Epidemiology: Integrating the Ideas, Theories, Principles and Methods of Epidemiology*, 3rd edn. New York: Oxford University Press, 2016.

22 Haynes RB, Sackett DL, Guyatt GH, Tugwell P. *Clinical Epidemiology: How to Do Clinical Practice Research*, 3rd edn. Philadelphia, PA: Lippincott-Raven, 2006.

23 Woodward M. *Epidemiology: Study Design and Data Analysis*, 3rd edn. Boca Raton, FL: Chapman & Hall/CRC, 2013.

24 Weiss NS, Koepsell TD. *Epidemiologic Methods: Studying the Occurrence of Illness*. New York: Oxford University Press, 2014.

25 Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; **32**: 51–63.

26 Mitchell AA, Cottler LB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.

27 Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; **58**: 295–300.

28 US Public Health Service. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC: Government Printing Office, 1964, p. 20.

29 Experimental carcinogenesis with tobacco smoke. In: *US Public Health Service: The Health Consequences of Smoking – Cancer. A Report of the Surgeon General*. Washington, DC: Government Printing Office, 1982, p. 181.

30 Biomedical evidence for determining causality. In: *US Public Health Service: The Health Consequences of Smoking – Cancer. A Report of the Surgeon General*. Washington, DC: Government Printing Office, 1982, p. 13.

31 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**: 878–81.

32 Joint Commission on Prescription Drug Use. *Final Report*. Washington, DC: Joint Commission on Prescription Drug Use, 1980.

33 Humphries TJ, Myerson RM, Gifford LM, *et al*. A unique postmarket outpatient surveillance program of cimetidine: report on phase II and final summary. *Am J Gastroenterol* 1984; **79**: 593–6.

34 Markush RE, Seigel DG. Oral contraceptives and mortality trends from thromboembolism in the United States. *Am J Public Health* 1969; **59**: 418–34.

35 National Center for Health Statistics. *Health: United States 1982*. Hyattsville, MD: Department of Health and Human Services, 1982.

36 US Public Health Service. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC: Government Printing Office, 1964, p. A1.

37 Strom BL, Stolley PD. Vascular and cardiac risks of steroidal contraception. In: Sciarra JW ed., *Gynecology and Obstetrics*, vol. **6**. Hagerstown, MD: Harper and Row, 1989, pp. 1–17.

38 Royal College of General Practitioners. *Oral Contraceptives and Health*. London: Pitman, 1974, ch. 7.

39 Vessey MP, Doll R, Peto R, Johnson B, Wiggins P. A long-term follow-up study of women using different methods of contraception – an interim report. *J Biosoc Sci* 1976; **8**: 373–427.

40 Bell RL, Smith EO. Clinical trials in post-marketing surveillance of drugs. *Control Clin Trials* 1982; **3**: 61–8.

41 Manson JE, Hsia J, Johnson KC, *et al*. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 2003; **349**: 523–34.

42 Herrington DM, Howard TD. From presumed benefit to potential harm: hormone therapy and heart disease. *N Engl J Med* 2003; **349**: 519–21.

43 John B. Hormone-replacement therapy and cardiovascular diseases. *N Engl J Med* 2003; **349**: 521–2.

44 Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002; **288**: 321–33.

45 Col NF, Pauker SG. The discrepancy between observational studies and randomized trials of menopausal hormone therapy: did expectations shape experience? *Ann Intern Med* 2003; **139**: 923–9.

46 Grimes DA, Lobo RA. Perspectives on the Women's Health Initiative trial of hormone replacement therapy. *Obstet Gynecol* 2002; **100**: 1344–53.

47 Whittemore AS, McGuire V. Observational studies and randomized trials of hormone replacement therapy: what can we learn from them? *Epidemiology* 2003; **14**: 8–9.

48 Mellin GW, Katzenstein M. The saga of thalidomide. *N Engl J Med* 1962; **267**: 1238–44.

49 Taussig HB. A study of the German outbreak of phocomelia. *JAMA* 1962; **180**: 1106–14.

50 Kajii T, Kida M, Takahashi K. The effect of thalidomide intake during 113 human pregnancies. *Teratology* 1973; **8**: 163–6.