SPECIAL ARTICLE

Methods in Epidemiology: Observational Study Designs

Natalie A. DiPietro, Pharm.D., M.P.H.

This article is the first of a three-part series intended to enhance clinical pharmacists' understanding of methods frequently used in epidemiologic research and their applications. The basic tenets of epidemiology and uses for data derived from epidemiologic studies are given, along with a high-level overview of the differences between experimental and observational study designs. The defining characteristics of each of the observational study designs (case report or case series, ecologic, cross-sectional, cohort, case-control, nested case-control, and case-cohort) and the resultant strengths and limitations of the study designs are presented. Applications for observational studies in pharmacoepidemiology (including the case-crossover and case-time-control study designs) are discussed. Finally, points to consider when evaluating data from observational studies are addressed.
Key Words: epidemiology, observational study, public health.
(Pharmacotherapy 2010;30(10):973–984)

Among practitioners, academia, and professional organizations in the United States, there is growing interest in roles and opportunities for pharmacists in public health.[1–9] To fully realize the applications for pharmacists in public health as well as in clinical practice, pharmacists must be knowledgeable about the tools and techniques used in public health research. Recent studies have shown that because of the widespread use of epidemiologic methods in published literature, pharmacists need to comprehend these types of study designs, along with the appropriate biostatistical tests, to effectively fulfill their professional responsibilities.[10, 11] Observational designs are regularly used in epidemiologic studies; with greater appreciation and insight into considerations in the design, analysis, and interpretation of data from these studies, pharmacists will be well positioned to make recommendations based on resultant data or to contribute to the design and implementation of such studies.

This is the first of a three-part series of articles intended to enhance clinical pharmacists' understanding of methods frequently used in epidemiologic research and their applications. The most common types of observational studies are highlighted in this article. The second article will review biostatistical techniques common in epidemiologic research, and the third article will cover drug literature evaluation for observational studies, including common biases that must be taken into consideration. Our intent is that after reading the three-part series, pharmacists will be able to better interpret literature from observa-

Table 1. Classification Scheme of Common Epidemiologic Study Designs

Type of Design	Investigators' Functions	Examples
Observational	Does not allocate exposure	
Descriptive	Characterize or compare disease frequency and pattern in populations	Case report, ecologic,
	Generate preliminary hypothesis of exposure-disease relationship	cross-sectional
Analytic	Gather more data and control for confounding factors	Cohort, case-control,
	Better establish cause-and-effect relationships	nested case-control,
		case-cohort
Experimental	Allocate exposure	Randomized clinical trial,
Interventional	Attempt to isolate effect of exposure only	community trial

tional studies and to better understand the relevant statistical analyses. Readers are encouraged to consult the publications cited in the reference section of each of the articles for more in-depth information on the proper design, implementation, and interpretation of observational study designs.

Epidemiologic Research Methods

Epidemiology is the science of the study of distributions and determinants of disease and wellness in populations.[12, 13] The pattern of disease distribution in populations is often described in the dimensions of person, place, and time.[14] Techniques to quantify the changing trends in disease incidence or prevalence are further described in the second article in this series. Epidemiologists are interested in determining associations between specific exposure(s) and specific outcome(s).[13] Exposures, such as lifestyle behaviors, biologic agents, or chemical agents, may increase the risk (risk factor) or decrease the risk (protective factor) of the development of a particular outcome (morbidity, mortality, or positive functioning).[13] Data derived from epidemiologic studies inform other aspects of public health (such as health promotion and health policy and planning) and clinical medicine.[15, 16]

Epidemiologists use both observational study designs and experimental study designs in their research (Table 1). The selection of the design to be used for a particular research question is dependent on a number of factors. The types of study designs available to researchers differ with respect to several dimensions including whether the outcome of interest is present or absent in study participants at the start of the study (directionality of exposure measurement); the unit (individual or group) and number (one or more) of observations made; and the method(s) and timing of data collection.[15, 17] Consideration

of these factors as well as ethical implications for the population to be studied[17, 18] and available resources should lead an investigator to select the best type of study design to address the question at hand.

Experimental and observational studies differ with regard to investigator manipulation of the exposure of interest.[17] In quasiexperimental or experimental study designs, the investigator allocates or controls the exposure of interest in an attempt to isolate the effect of the exposure only; therefore, causal associations can be better established.[17] Depending on the research question, an interventional trial may be the best option for epidemiologists to use. One example is the Multiple Risk Factor Intervention Trial (MRFIT) in which 12,866 men aged 35–57 years at high risk for coronary heart disease were randomly assigned to receive either a program of counseling on diet and smoking cessation and hypertension treatment (according to protocol) or usual care from community sources.[19] Another type of experimental trial design of use to epidemiologists is known as a community trial, where an entire community or population (such as a town, school, or work site) is assigned to a particular intervention and is compared with a similar population that did not have the exposure.[18] An example of a community trial would be to compare the rate of dental caries in a community with a fluoridated public water supply with the rate in a community that does not add fluoride to the water.[18]

The randomized controlled trial is the most scientifically rigorous method for hypothesis testing.[17, 20] The validity of the inference of the association between exposure and outcome is further enhanced if the study is well designed and executed.[17, 18, 20] However, investigators may be limited in their ability to use an experimental design because of issues related to feasibility, ethics, or cost.[17, 20]

In an observational study, the investigator does

Table 2. Differentiating Characteristics of Observational Study Designs

Study Design	Level of Observation	Patients Identified Based on Exposure or Outcome Status	Measure of Association	Major Design Advantages	Major Design Disadvantages
Case report or case series	Individual	Outcome	Usually cannot be calculated	Generate information on natural history of disease and case definition Identify new disease or condition	Usually cannot calculate rates or measures of association
Ecologic	Group	Not applicable	Correlation	Relatively quick and inexpensive	Ecologic fallacy
Cross-sectional	Individual	Not applicable	Prevalence	Relatively quick and inexpensive	Temporal association cannot be established
Case-control	Individual	Outcome	Odds ratio	Ability to study relatively rare outcomes Less expensive and faster than cohort studies (especially prospective cohort studies)	Not practical for study of rare exposure Inability to study multiple outcomes in one study
Cohort	Individual	Exposure	Relative risk	Ability to study relatively rare exposures Temporal association most clearly delineated Can calculate direct estimates of risk and incidence	Not practical for study of rare outcomes Increased cost and length (especially prospective design)

not allocate or control the exposure; rather, the investigator observes and evaluates the results that occur without intervention.[17] Observational study designs include the case report and case series, and ecologic, cross-sectional, cohort, case-control, nested case-control, and case-cohort studies. Table 2 provides a high-level overview of the distinguishing characteristics of these observational study designs. Two designs also used (often in pharmacoepidemiologic research) are the case-crossover study and case-time-control study. Each observational study design and its important differentiating characteristics are described in the remainder of this article. It is important to note that each type of study has inherent strengths and limitations due to its design elements. Although observational study designs have been traditionally classified as either descriptive or analytic based on these strengths and limitations (Table 1), this may be regarded as an oversimplification.[15, 18] Each of these epidemiologic study designs at its core intends to examine person, place, and time and to contribute to the understanding of the etiologic basis of disease.[15] However, the design elements of some of the studies may limit the investigators' ability to thoroughly and rigorously test the cause-and-effect hypothesis.[17]

Epidemiologists may use primary data collected specifically for the study at hand (referred to as primary data), preexisting secondary data (such as physician or hospital records, vital statistics data, or insurance databases), or a combination of both in their research.[21] Although some types of observational studies use surveys as a tool to gather primary data, a complete discussion of survey research and the methodology of designing a good survey is outside the scope of this article. The reader is referred to a brief list of excellent resources on this topic (Appendix 1).

Descriptive Observational Study Designs

Descriptive study designs include the case report or case series, ecologic study, and cross-sectional study.[17] Through the use of descriptive study designs, investigators can establish disease frequency and pattern in a population and provide data to be used not only for health policy and planning, but also as preliminary information to consider possible associations between exposure and outcome.[17] Due to intrinsic limitations that are described later, these studies are sometimes called "hypothesis generating,"[17, 18, 22] as often the proposed relationship between exposure and outcome will need to be studied further with an analytic design[22] to allow for

more complete data gathering and sophisticated statistical analyses.

Case Report or Case Series

Clinical case reports or case series involve the publication of one or a number of interesting clinical cases observed by a health care professional or a group of health care professionals, usually in a specific geographic region.[15] Clinical case series are helpful to epidemiologists and clinicians alike, as the descriptions of unique cases help to generate a case definition for disease as well as information on the natural history of disease, including the range of signs, symptoms, and consequences of the disease that may manifest in an individual.[15] In addition, cases can be analyzed to identify trends or similarities that may provide a basis for research into possible etiologic factors of the disease; for example, case reports generated the hypothesis that rubella during pregnancy may result in congenital cataracts.[15]

Case series are also useful in clinical teaching.[15] New diseases or epidemics may be first identified through clinical case series; Legionnaire's disease and acquired immunodeficiency syndrome were two diseases first described in case series.[15] Generally, measures such as incidence or prevalence rates cannot be calculated, as complete counts of all cases and/or the population at risk are usually not available.[15] However, if a clinical case series contains a complete count of the cases identified in a specific geographic area for a known, defined population, it may be possible to calculate incidence or prevalence rates.[15]

Ecologic Study

Ecologic studies examine data for groups (such as communities, countries, or other aggregate units or population levels) to identify possible associations between exposure and outcome.[15, 17, 18] In an ecologic comparison study (also known as cross-sectional ecologic study), investigators examine the relationship between exposure rates and disease rates in populations in a certain period of time.[17] An article published in 2009 provides an example of an ecologic study.[23] In this study, investigators obtained data on road deaths and injuries from the Fatality Analysis Reporting System and examined the impact of increased speed limits in the United States on fatalities and injuries from fatal crashes. Another example is a study conducted in England that examined a possible association between administration of the measles-mumps-rubella (MMR) vaccine and the subsequent development of Crohn's disease.[24] The investigator used hospital admissions data to identify whether there were changes in age-specific rates of Crohn's disease among children and adolescents after the introduction of the MMR vaccine in 1988. Ecologic trend studies are ecologic studies performed within the same population at different time points to assess changes over time.[17] The measure of association for an ecologic study is correlation.[17]

Advantages of ecologic studies include minimal cost and duration compared with those of other types of study designs.[17] Ecologic studies have utility in hypothesis generation and preliminary research, especially for areas in which there are little existing data regarding the potential association between an exposure and an outcome.[17] A major limitation of this study design is "ecologic fallacy," a term used to represent the fact that associations observed at the level of the group or population may not represent the exposure-disease association at the individual level.[17] In an ecologic study, data are available on the exposure and outcome of interest at a population level, but not every individual in the population studied may have had the exposure of interest or the same amount of that exposure.[15, 17] Because of the nature of the data analyzed in an ecologic study, it is not possible to link exposure with outcome in particular individuals within the population[25]; therefore, with the data available, it is not possible to assess whether individuals with the exposure of interest developed the outcome of interest.

Migration of individuals within populations may further complicate conclusions from ecologic studies.[17] The data used in ecologic studies to assess exposures or risk factors (such as data on taxes collected through tobacco sales or gun purchases) may be deficient markers for the true exposure of interest.[18] Finally, it is important to remember that correlation does not necessarily substantiate causation; just because a correlation between the exposure and outcome is found, a causal association cannot be established from ecologic studies.[17]

Cross-sectional Study

The cross-sectional study, also sometimes called prevalence study, is another type of descriptive study design available to epidemiologists. Cross-sectional studies can be useful

when investigators are interested in gathering information on risk factors and extent of disease in a particular population or in characterizing or comparing populations.[15] In this study design, investigators select a sample of subjects and collect data on health status. Usually, information on both exposures and disease are collected in a single visit[17, 18]; however, in some studies investigators may choose to collect data only on exposure or only on disease.[17]

Examples of cross-sectional studies include the National Health and Nutrition Examination Survey (NHANES) and the Behavioral Risk Factor Surveillance System (BRFSS), both conducted by the Centers for Disease Control and Prevention. Through NHANES, a representative sample of approximately 5000 individuals was surveyed across the United States to obtain information on health and nutritional status.[26] The BRFSS is administered by individual states to assess adults' health-related behaviors, both risky and protective, that may influence their health status.[27] Both surveys provide timely data that may inform health policy and planning.[26, 27] In an example of a cross-sectional study conducted by pharmacists in a community setting, the investigators surveyed individuals attending a health screening fair to assess their knowledge of risk factors for cardiovascular disease.[28]

Cross-sectional studies are typically faster and less expensive to conduct than analytic study designs.[17] Although cross-sectional studies are valuable in that investigators are able to estimate the prevalence of disease as well as risk factors and exposures, a major limitation is that a temporal association cannot be established.[17] Since data on exposure and outcome are obtained at the same time, investigators are not able to establish with certainty that the exposure preceded the outcome and should not infer cause and effect based on the data collected during a cross-sectional study.[17] Other limitations of cross-sectional studies include limited utility in studying rare diseases or diseases with short duration.[17] In addition, survivor bias may be introduced as the most severe or fatal cases may be excluded from the study due to the sampling frame and/or methodology used in a particular study.[15, 17]

In most cross-sectional studies, investigators do not choose subjects based on exposure or disease status; optimally, all individuals in a defined geography and time are studied or a random, representative sample of such population is selected.[15, 18] However, investigators

sometimes elect to study a convenience sample, which may limit the reliability and generalizability of study results.[17] Within the sample, investigators may choose to compare the risk factors and outcomes reported in various study subgroups (e.g., men vs women or employed vs unemployed).[15] Depending on the variables of interest, investigators may also choose to collect data over a relatively long period of time such as 1 year or repeat the cross-sectional study at different times of the year to account for seasonal variations.[15] Cross-sectional studies may also be repeated over time to allow for study of changing trends in disease or risk factors.[17] A variation of the cross-sectional study, sometimes referred to as a prevalent case-control study, involves differential sampling of subjects due to disease status; the design may be similar to that of a case-control study (described below) except that prevalent cases, rather than incident case, are included.[18]

Applications of the cross-sectional study are numerous. Data gathered on exposures and outcomes can generate possible hypotheses to study in more rigorous, analytic studies.[17] In addition, understanding distributions of disease and exposure in populations allows clinicians, public health professionals, and health policy makers to design and implement appropriate interventions or allocate resources effectively.[17]

Analytic Observational Study Designs

Analytic studies include cohort, case-control, nested case-control, and case-cohort.[17] Analytic designs typically involve more thorough data collection and analyses than the descriptive studies described above; therefore, investigators using these designs can better examine possible associations between exposure and outcome and identify potential preventive measures.[17, 18] Cause-and-effect inferences can be better established due to the characteristics of the analytic designs in which investigators may analyze and control for extraneous factors that may influence the observed association between exposure and outcome (known as confounding factors) and better quantify the temporal sequence of exposure and outcome.[18]

Cohort Study

The cohort study is considered to be one of the most powerful types of observational studies.[29] Investigators initiating a cohort study define the population groups (or subsets) to be examined

based on the presence or absence of a particular exposure of interest. At baseline, an individual's exposure status is ascertained.[29] The members of the cohort are then followed over a period of time to observe the occurrence of the outcome(s) of interest among the exposed and nonexposed groups.[29] Investigators can design cohort studies as prospective, retrospective, or ambispective.

As the name suggests, a prospective cohort study follows individuals from baseline to a point in the future.[29] Well-known examples of prospective cohort designs include the Nurses' Health Study and the Framingham Heart Study. Since the 1970s, the Nurses' Health Study has obtained data from more than 230,000 female nurses by using mailed surveys.[30] The surveys, which collect data on a range of health-related topics and behaviors such as smoking status, nutritional status, and the development of specific diseases, have been mailed to participants periodically (every 2–4 yrs, based on the specific questionnaire).[30] The third cohort study, currently being designed, will be entirely Web based, and the investigators plan to recruit women of more diverse race-ethnicity than was studied in the first two cohorts.[30] The Framingham Heart Study was started in 1948 and followed more than 5000 men and women in the town of Framingham, Massachusetts, to examine the determinants of cardiovascular disease; the second and third generations of this population are now being studied.[31] Over 1900 publications have been generated by the data from the Framingham Heart Study between 1950 and 2008.[31]

A retrospective cohort study makes use of historical data to reconstruct an individual's past exposure status at baseline and subsequent outcomes that have occurred and been recorded before the study was started.[29, 32] One example is a study published in 2009 where the investigators performed a retrospective analysis for 8205 patients with acute coronary syndrome from 127 Veterans Affairs hospitals to examine outcomes in individuals taking clopidogrel with or without a proton pump inhibitor.[33]

The ambispective cohort design is a blend of the retrospective and prospective designs; retrospective data are used to determine exposure status, and participants are then followed into the future to obtain information on disease status.[29] Retrospective and ambispective study designs require solid historical data to be effective.[29, 32]

Regardless of the study design (prospective, retrospective, or ambispective), cohort studies are future oriented with regard to the natural history of a disease,[15] as the study subjects are free of disease at the start of the study.[29] In each of the cohort study designs, an individual's exposure status is measured before the outcome is recognized or diagnosed.[32] Thus, the association obtained from a cohort study has traditionally been considered the strongest evidence among observational study designs, as the temporal association between exposure and disease can be clearly delineated and it can be verified that exposure preceded outcome (especially in prospective studies).[29] The design of the cohort study allows investigators to calculate incidence rates[31]; the measurement of association obtained in a cohort study is the relative risk.[29] Other advantages of cohort studies include the ability of investigators to study the effects of rare exposures[29] and to study the various outcomes or effects of a single exposure.[29, 34] Cohort studies can also provide information on the natural history of a disease, including a determination of the time between exposure and development of disease.[29]

Disadvantages to the use of the cohort designs, especially the prospective cohort design, include more time and expense required than for other types of observational studies.[29] Cohort studies are typically not practical for studies of rare diseases as the sample size needed to detect such outcomes would be too large.[29, 35] A bias specific to cohort studies is immortal person-time.[35] If the inclusion criteria for a cohort study require exposure for a certain period of time before enrollment in the cohort, then only those who have "survived" the exposure will be included for analysis.[35] Cohort studies may be susceptible to loss to follow-up bias.[29] (The third article in this series will further discuss this issue.) In addition, it is important to be aware of the potential for misclassification of exposure if during a long cohort study an individual's exposure changes or if an individual's exposure status changes as a result of being included in the study.[29] Finally, ethical issues may arise in a prospective cohort study, for it may be unethical for investigators not to intervene if certain exposures have already been shown to have negative effects.[29]

Case-Control Study

The case-control study is a type of analytic observational study design in which investigators select participants based on disease status.[17] Case-control studies are largely retrospective,[15, 17, 36] and

data on disease and exposure are collected at a single point in time.[17] Information on past exposures informs investigators on potential etiologic factors associated with the disease of interest.[17] Cases are individuals who have the disease of interest; controls are individuals who do not have the disease of interest.[17] An example of a case-control study was a study conducted by using the United Kingdom General Practice Research Database (UKGPRD), a database to which certain physicians in the United Kingdom contribute data regarding their patients.[37] In this study, cases (defined as individuals born in or after 1973 with a diagnosis of pervasive developmental disorder between 1987 and 2001) were identified from the UKGPRD. Controls were matched to cases by age, sex, and practice of physician. The investigators then looked at medical records to assess an individual's MMR vaccination status.[37]

Proper and unbiased selection of cases and controls is critical to the validity of a case-control study.[36] Sometimes, investigators choose to match cases and controls based on certain characteristics in order to control for confounding.[38] Much has been published on the nuances of case and control selection and matching; key points are summarized here, but the reader is referred to other sources for more detailed information.[15, 17, 39–43]

A detailed case definition should be used to facilitate the identification of cases.[17] The ideal case definition will yield representative and incident (or early diagnosis) true cases of the disease of interest from a specific population for a specific period of time.[15, 17] Investigators must find the right criteria to avoid misclassification; however, criteria that are too restrictive may not be feasible as enrollment could be severely limited.[17] Ideally, all false cases would be excluded and all true cases of the disease would have an equal probability of being placed in the study.[17] Investigators should identify controls from the population at risk for the outcome being studied in the cases.[15, 17] A commonly cited litmus test for proper control selection is to assess whether all controls would have been included in the study as a case if they had developed the outcome of interest.[17] Sources of controls include population-based, hospital-based, or acquaintance-based (friends or relatives of the cases) sources.[17] Each of these potential sources of controls has strengths and advantages, but each source also introduces the potential for a different form of bias.[17] These biases include possible responder

bias (population-based), Berkson's bias (hospital-based), or overmatching (acquaintance-based)[17]; these will be discussed in more detail in the third article.

The measure of association obtained from a case-control study is an odds ratio.[17] Incidence rates, and therefore, risk ratios, cannot be calculated from a case-control study.[17] However, the odds ratio can closely approximate the relative risk if certain assumptions are met: representative cases (with regard to severity and case definition) were selected; unbiased, representative controls (with regard to the frequency to exposure of interest) were selected; and the disease of interest is rare (although not all experts in the field of epidemiology agree that the third assumption is necessary).[15, 17]

Since participants are selected on outcome status, advantages of the case-control design include the ability to study relatively rare outcomes or diseases, or such conditions with long latency periods.[17] Case-control studies also afford the opportunity to study multiple exposure(s).[36] Case-control studies are typically less expensive and quicker to complete than cohort studies.[17] Limitations of this study design may include possible biases such as temporal bias (the inability to ensure that exposure preceded disease), selection bias, and recall bias (see third article of this series).[17, 36] Case-control studies are neither suitable to study rare exposures nor to study more than one disease state or outcome in one study.[17]

Nested Case-Control and Case-Cohort Studies

Two study designs that incorporate features of both the cohort study (prospective or retrospective) and the case-control study are the nested case-control and case-cohort studies. A nested case-control study is a case-control study in which cases and controls are selected from among members of a cohort.[32, 41] For example, cases with the disease of interest that arises during the course of the cohort study may be compared with a subset of individuals from the population that did not develop the disease of interest (controls).[32, 41] A control may later become a case; however, this does not typically occur when the disease of interest is uncommon.[32] Cases are usually matched by certain variables such as sex, age, and time of enrollment into the cohort.[32] An example of a nested case-control study examined whether there was an association between exposure to asbestos and development of

autoimmune disease.[44] From a cohort of over 7000 current and former residents of Libby, Montana (a location with a history of occupational and community exposure to asbestos), 494 potential cases (individuals with a diagnosis of one of three autoimmune diseases under study) were selected. Controls (1482 subjects) were selected from the remaining cohort in a 3:1 ratio and frequency-matched on age and sex groupings.

Nested case-control studies are analyzed with similar techniques as other case-control studies.[32] A major advantage of this study design is that since baseline data were obtained on exposure status and other clinical attributes, certain biases such as recall bias and selection bias may be minimized.[45] In addition, if certain data were collected from all participants at baseline (e.g., blood samples), analysis of biomarkers or other cost-intensive tests can be carried out in a subset of individuals, not the entire cohort.[29, 32, 41] The design of a nested case-control study allows for the estimation of absolute and relative risk functions.[46] Limitations with this study design may arise; for example, if controls are sampled at the end of the study period, issues such as loss to follow-up may come into play.[43] An additional limitation of obtaining controls at the end of the study period is that those controls may not be representative of all controls if the outcome of interest is not rare.[43]

In a case-cohort study (also called a case-base study), investigators select cases from the members of the cohort who have the outcome(s) of interest and select a sample of controls (sometimes called a subcohort) randomly from the rest of the same cohort.[32] Unlike a nested case-control design where controls are usually matched to cases on time of entry into the cohort, in a case-cohort study every individual in the cohort has an equal probability of being a control since the control groups are randomly selected.[32, 41] Similar to a nested-case control study, the design of a case-cohort study permits investigators to obtain certain information for only a subset of all controls, potentially saving time and money[32] and minimizing certain biases.[45] The group of individuals serving as controls may be used for multiple case groups.[32] Controls that later develop the disease(s) of interest may later enter the study as a case[32, 41]; therefore, investigators must select more controls for each case than they would have to in a traditional case-control study to attain the same level of statistical precision.[41] Loss to follow-up, frequency of the outcome of interest, and the possibility of the exposure changing over time may limit the feasibility of using the case-cohort design.[43]

An example of a case-cohort study was a study published in 2006 in which the purpose of the study was to examine the possible association between iron levels and coronary heart disease or risk of acute myocardial infarction.[47] Investigators compared four biomarkers for blood iron levels among all 165 members of the cohort who developed coronary heart disease and a random sample of 1134 controls from the baseline cohort of 11,471 postmenopausal women enrolled in the Prospect-EPIC (European Prospective Investigation into Cancer and Nutrition) study.

Measures of association generated from a case-cohort study include rate ratio, risk ratio, and odds ratio.[32] Investigators may choose to compare the measures of association calculated from the case-cohort study to measures of association in populations external to the study.[32]

Applications in Pharmacoepidemiology

Observational study methods are often used in pharmacoepidemiology, which is the study of drug use, efficacy, and toxicity in populations.[48, 49] Observational study designs tend to be important mechanisms for the identification and evaluation of adverse drug events. First, observational studies permit the examination of unintended negative effects that would not be ethical to study through experimental trials. Second, most clinical studies evaluating a drug's safety and efficacy are conducted in very controlled clinical settings and with (relatively) limited numbers of patients for a specified period of time. The observational studies that can be performed once a drug is on the market can allow for the examination of adverse events that occur in a more diverse or special (e.g., pediatric, pregnant) patient population, that are uncommon or latent, that are the result of a drug overdose, or that result from off-label use of the drug product.[48, 50]

Each of the study methods described above can be applied in pharmacoepidemiologic research,[51] although limitations and biases may be introduced[49, 52, 53] (a thorough discussion of which is outside the scope of this article). Additional types of observational study designs that are well suited to answer questions in pharmacoepidemiology include the case-crossover study and the case-time-control study. (Note that these study designs may be used to answer other types of epidemiologic questions as well.)

Case-Crossover Study

The design of the case-crossover study is similar to the design of the experimental crossover study in that the same individual is assessed during periods of a specific drug exposure and periods without that exposure (control period).[49, 54] In this setting, the case has the outcome of interest and serves as its own control.[45, 54] The design is particularly useful to examine the effects of drug use in patients with diseases that worsen over time or that vary in severity from patient to patient or the effects of intermittent drug use.[49] The outcome of interest must be a discrete event, and a risk period (between exposure and outcome) should be specified.[49, 54] Limitations of the case-crossover study include recall bias[45, 49] and logistical issues of implementing such a design with a secondary database, such as claims data. As the drug studied in this type of design is typically not used regularly, it can be difficult to ascertain when the drug was actually taken (vs when it was dispensed).[49] Additional assumptions with this type of study design are as follows: first, the effect of the exposure is not cumulative or does not extend beyond the risk period, and second, the outcome of interest is without a preclinical stage that may influence the exposure.[43, 45]

Case-Time-Control Study

The case-time-control study design is an extension of the case-crossover design and is used to examine associations that may exist between exposure and outcome in situations where trends that may change over time (such as prescribing patterns or disease severity) could confound the association.[49, 54] A control group is selected from among the cohort under study with an approximate synchronization with cases.[49, 52] Both cases and controls are examined for exposure status during the control period and during the time period corresponding to the outcome of interest.[52] As cases and controls were selected from approximately the same time period, changes in trends over time may be adjusted for, although this is not guaranteed and may itself introduce other bias.[52]

Evaluating Data from Observational Studies

Observational studies, like experimental trials, are susceptible to error, bias, and confounding that may lead to erroneous finding(s) and/or conclusion(s) in the study (see third article in this series). Critical evaluation must take into account some issues that are distinctive to observational study designs. Observational studies differ from experimental studies in that when using observational study designs, the investigator does not manipulate the exposure of interest. As a result, observational studies have some limitations relative to experimental studies; due to the lack of randomization, investigators cannot fully control for residual (unknown) confounding in observational studies.[18, 55] For a brief illustrative example, consider the discrepant data from observational studies and randomized clinical trials regarding the effect of postmenopausal hormone replacement therapy on certain end points (i.e., heart disease). Limitations in the design and implementation of the observational and experimental studies were considered as possible explanations for the discrepant data.[50] In addition, it is important to acknowledge that residual confounding in the observational studies may have affected the observed association between exposure and outcome.[55] The act of randomization, if conducted in a proper and unbiased fashion, should control for this residual confounding.[18] However, randomized trials can still be susceptible to bias if there are deficiencies in design or data analysis.[50]

Even with the concern of residual confounding, the data obtained from well-designed observational studies contribute to the body of evidence and overall scientific understanding of a particular exposure-outcome association. Specific factors have been established as points of consideration when assessing whether the observed association is in line with a cause-and-effect relationship.[16, 51, 56] Although not without limitations,[57, 58] this framework is often referenced in epidemiologic texts. Points to consider include strength of the association,[16, 51, 56] consistency on repetition,[16, 51, 56, 57] specificity,[16, 51, 57] time sequence[16, 51, 56, 57] dose-response relationship (or biologic gradient),[16, 51, 56, 57] biologic plausibility,[16, 51, 56, 57] coherence of explanation,[16, 51, 57] and analogy.[16, 57] Table 3 provides a brief description of each of these parameters.[16, 51, 56–58] Although not meant to be used as the lone criterion for causal inference,[16, 57] researchers can consider each point when studying a particular association.[56] The more of these parameters satisfied by the observed association, the greater the likelihood that the observed association may be causal.[16, 51] Through the use of varied study designs, epidemiologists strive to hypothesize and analyze cause-and-

Table 3. Points to Consider When Assessing a Possible Association Between Exposure and Outcome

Parameter	Brief Description	Comments
Strength of the association	Magnitude of observed association between exposure and outcome	The stronger the association, the less likely it is due to an uncontrolled or unknown confounding variable; however, just because the magnitude of an association is not high does not preclude a cause-and-effect relationship
Consistency on repetition	Reproducibility; whether the same association has been observed by different persons in different settings	May be especially important when examining rare exposures
Specificity	Whether the exposure occurs without the subsequent outcome of interest, or whether the outcome occurs without the exposure of interest	Rarely met; however, if the observed association meets this consideration, this is strong evidence for a cause-and-effect relationship
Time sequence	Ensuring exposure occurred before outcome or disease of interest and that time period between exposure and outcome has biologic plausibility	Not all observational study designs allow for study of temporal sequence
Dose-response relationship	Gradient of risk is associated with the degree of exposure	Absence of dose-response relationship does not preclude cause-and-effect relationship; all cause-and-effect relationships do not have a dose-response relationship or may have a threshold phenomenon
Biologic plausibility	Known or hypothesized biologic mechanism by which the exposure may change risk of developing the outcome	Knowledge changes over time, affecting perception of biologic plausibility
Coherence of explanation	Whether the association contradicts what is already known about the biology or natural history of the disease of interest	Conflicting information may be incorrect or misinterpreted
Analogy	Whether there are other associations between exposure and outcome similar to the one under study	May conflict with other considerations such as specificity

effect relationships with the intent that the data can be used to ultimately prevent disease and improve public health.[58]

Conclusion

The focus of epidemiology is to examine distributions and determinants of disease and wellness in populations. Data derived from epidemiologic studies inform clinical medicine and public health; epidemiologic techniques are also applied in pharmacoepidemiologic research. The various study designs used in epidemiology have inherent strengths and limitations that affect the extent to which a causal association between exposure and outcome can be inferred. All types of studies may be affected by error, bias, or confounding. Clinical pharmacists must critically evaluate and interpret data from epidemiologic studies to effectively fulfill their professional responsibilities.

References

1. **Calis KA, Hutchison LC, Elliott ME, et al, for the American College of Clinical Pharmacy**. Healthy people 2010: challenges, opportunities, and a call to action for America's pharmacists. Pharmacotherapy 2004;24:1241–94.
2. **Babb VJ, Babb J**. Pharmacist involvement in healthy people 2010. J Am Pharm Assoc 2003;43:56–60.
3. **Joint Commission on Pharmacy Practice**. Future vision of pharmacy practice 2015. Available from http://www.pharmacist.com/. Accessed February 15, 2009.
4. **Healthy People Curriculum Task Force**. Clinical prevention and population health curriculum framework. Available from http://www.atpm.org/resources/pdfs/Official_CPPH_Framework_2008.pdf. Accessed November 25, 2008.
5. **Center for the Advancement of Pharmaceutical Education**. Educational outcomes 2004. Available from http://www.aacp.org/Docs/MainNavigation/Resources/6075_CAPE2004.pdf. Accessed November 25, 2008.
6. **Accreditation Council for Pharmacy Education**. Accreditation standards and guidelines for the professional program in pharmacy leading to the doctor of pharmacy degree (adopted: January 15, 2006; effective: July 1, 2007). Available from http://www.acpe-accredit.org/pdf/ACPE_Revised_PharmD_Standards_Adopted_Jan152006.pdf. Accessed November 25, 2008.
7. **American Public Health Association**. The role of the pharmacist in public health. Available from http://www.apha.org/advocacy/policy/policysearch/default.htm?id=1338. Accessed November 25, 2008.
8. **American Pharmacists Association**. The role of pharmacists in public health awareness. Adopted policy statements 1963–2007. Washington, DC: American Pharmacists Association, 2007:41.
9. **American Society of Health-System Pharmacists**. Statement on the role of health-system pharmacists in public health. Am J Health Syst Pharm 2008;65:462–7.

10. **Draugalis JR, Plaza CM**. Emerging role of epidemiologic literacy. Ann Pharmacother 2006;40:229–33.
11. **Lee CM, Soin HK, Einarson TR**. Statistics in the pharmacy literature. Ann Pharmacother 2004;38:1412–18.
12. **Hennekens CH, Buring JE**. Definition and background. In: Mayrent SL, ed. Epidemiology in medicine. Philadelphia, PA: Lippincott Williams & Wilkins, 1987:3–15.
13. **Friis RH, Sellers TA, eds**. The history and scope of epidemiology. In: Epidemiology for public health practice, 3rd ed. Sudbury, MA: Jones and Bartlett, 2004:1–40.
14. **Friis RH, Sellers TA, eds**. Descriptive epidemiology: person, place, and time. In: Epidemiology for public health practice, 3rd ed. Sudbury, MA: Jones and Bartlett, 2004:127–69.
15. **Bhopal R, ed**. Epidemiological study design and principles of data analysis: an integrated suite of methods. In: Concepts of epidemiology, 2nd ed. New York, NY: Oxford University Press, 2008:285–346.
16. **Friis RH, Sellers TA, eds**. Practical applications in epidemiology. In: Epidemiology for public health practice, 3rd ed. Sudbury, MA: Jones and Bartlett, 2004:45–81.
17. **Friis RH, Sellers TA, eds**. Study designs: ecologic, cross-sectional, case-control. In: Epidemiology for public health practice, 3rd ed. Sudbury, MA: Jones and Bartlett, 2004:213–47.
18. **Rothman KJ, Greenland S, Lash TL**. Types of epidemiologic studies. In: Rothman KJ, Greenland S, Lash TL, eds. Modern epidemiology. Philadelphia, PA: Lippincott Williams & Wilkins, 2008:87–99.
19. **Multiple Risk Factor Intervention Trial Research Group**. Multiple risk factor intervention trial: risk factor changes and mortality results. J Am Med Assoc 1982;248:1465–77.
20. **Hennekens CH, Buring JE**. Intervention studies. In: Mayrent SL, ed. Epidemiology in medicine. Philadelphia, PA: Lippincott Williams & Wilkins, 1987:178–214.
21. **Mann CJ**. Observational research methods. Research design II: cohort, cross-sectional, and case-control studies. Emerg Med J 2003;20:54–60.
22. **Hennekens CH, Buring JE**. Design strategies in epidemiologic research. In: Mayrent SL, ed. Epidemiology in medicine. Philadelphia, PA: Lippincott Williams & Wilkins, 1987:16–29.
23. **Friedman LS, Hedeker D, Richter ED**. Long-term effects of repealing the national maximum speed limit in the United States. Am J Public Health 2009;99:1626–31.
24. **Seagroatt V**. MMR vaccine and Crohn's disease: ecological study of hospital admissions in England, 1991 to 2002. Br Med J 2005;330:1120–1.
25. **Hennekens CH, Buring JE**. Descriptive studies. In: Mayrent SL, ed. Epidemiology in medicine. Philadelphia, PA: Lippincott Williams & Wilkins, 1987:101–31.
26. **Centers for Disease Control and Prevention**. About the national health and nutrition examination survey, July 2009. Available from http://www.cdc.gov/nchs/nhanes/about_nhanes. htm. Accessed November 7, 2009.
27. **Centers for Disease Control and Prevention**. About the BRFSS, November 2008. http://www.cdc.gov/brfss/about.htm. Accessed November 7, 2009.
28. **Mooney LA, Franks AM**. Evaluation of community health screening participants' knowledge of cardiovascular risk factors. J Am Pharm Assoc 2009;49:529–37.
29. **Friis RH, Sellers TA, eds**. Study designs: cohort studies. In: Epidemiology for public health practice, 3rd ed. Sudbury, MA: Jones and Bartlett, 2004:253–87.
30. **Nurses' Health Study Group**. The nurses' health study, 2008. Available from http://www.channing.harvard.edu/nhs/. Accessed November 7, 2009.
31. **National Heart, Lung, and Blood Institute and Boston University**. Framingham heart study, September 2009. Available from http://www.framinghamheartstudy.org/. Accessed November 7, 2009.
32. **Kelsey JL, Whittemore AS, Evans AS, Thompson WD, eds**. Retrospective cohort, nested case-control, and case-cohort studies: planning and execution. In: Methods in observational epidemiology, 2nd ed. New York, NY: Oxford University Press, 1996:113–29.
33. **Ho PM, Maddox TM, Wang Li, et al**. Risk of adverse outcomes associated with concomitant use of clopidogrel and proton pump inhibitors following acute coronary syndrome. JAMA 2009;301:937–44.
34. **Hennekens CH, Buring JE**. Cohort studies. In: Mayrent SL, ed. Epidemiology in medicine. Philadelphia, PA: Lippincott Williams & Wilkins, 1987:153–77.
35. **Rothman KJ, Greenland S, Lash TL, eds**. Cohort studies. In: Modern epidemiology. Philadelphia, PA: Lippincott Williams & Wilkins, 2008:100–10.
36. **Hennekens CH, Buring JE**. Case-control studies. In: Mayrent SL, ed. Epidemiology in medicine. Philadelphia, PA: Lippincott Williams & Wilkins, 1987:132–52.
37. **Smeeth L, Cook C, Fombonne E, et al**. MMR vaccination and pervasive developmental disorders: a case-control study. Lancet 2004;364:963–9.
38. **Hennekens CH, Buring JE**. Analysis of epidemiologic studies: evaluating the role of confounding. In: Mayrent SL, ed. Epidemiology in medicine. Philadelphia, PA: Lippincott Williams & Wilkins, 1987:287–326.
39. **Kelsey JL, Whittemore AS, Evans AS, Thompson WD, eds**. Case-control studies. I. Planning and execution. In: Methods in observational epidemiology, 2nd ed. New York, NY: Oxford University Press, 1996:188–213.
40. **Kelsey JL, Whittemore AS, Evans AS, Thompson WD, eds**. Case-control studies. II. Further design considerations and analysis. In: Methods in observational epidemiology, 2nd ed. New York, NY: Oxford University Press, 1996:214–43.
41. **Rothman KJ, Greenland S, Lash TL, eds**. Case-control studies. In: Modern epidemiology. Philadelphia, PA: Lippincott Williams & Wilkins, 2008:111–27.
42. **Rothman KJ, Greenland S, Lash TL, eds**. Design strategies to improve study accuracy. In: Modern epidemiology. Philadelphia, PA: Lippincott Williams & Wilkins, 2008:168–82.
43. **Grobbee DE, Hoes AW, eds**. Case-control studies. In: Clinical epidemiology. Sudbury, MA: Jones and Bartlett, 2009:226–69.
44. **Noonan CW, Pfau JC, Larson TC, Spence MR**. Nested case-control study of autoimmune disease in an asbestos-exposed population. Environ Health Perspect 2006;114:1243–7.
45. **Szklo M, Nieto FJ, eds**. Basic study designs in analytical epidemiology. In: Epidemiology: beyond the basics. Sudbury, MA: Jones and Bartlett, 2004:3–52.
46. **Breslow NE**. Statistics in epidemiology: the case-control study. J Am Stat Assoc 1996;91:14–28.
47. **van der A DL, Marx JJM, Grobbee DE, et al**. Non–transferrin-bound iron and risk of coronary heart disease in post-menopausal women. Circulation 2006;113:1942–9.
48. **Strom BL, ed**. What is pharmacoepidemiology? In: Pharmaco-epidemiology, 4th ed. Hoboken, NJ: Wiley, 2005:3–16.
49. **Etminan M, Samii A**. Pharmacoepidemiology. I. A review of pharmacoepidemiologic study designs. Pharmacotherapy 2004;24:964–9.
50. **Etminan M, Gill S, FitzGerald M, Samii A**. Challenges and opportunities for pharmacoepidemiology in drug-therapy decision making. J Clin Pharmacol 2006;46:6–9.
51. **Strom BL, ed**. Study designs available for pharmacoepidemiology studies. In: Pharmacoepidemiology, 4th ed. Hoboken, NJ: Wiley, 2005:17–28.
52. **Schneeweiss S, Sturmer T, Maclure M**. Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. Pharmacoepidemiol Drug Saf 1997;6(S3):S51–9.
53. **Csizmadi I, Collet JP, Boivin, FJ**. Bias and confounding in pharmacoepidemiology. In: Strom BL, ed. Pharmacoepidemiology, 4th ed. Hoboken, NJ: Wiley, 2005:791–810.
54. **Donnan PT, Wang J**. The case-crossover and case-time-control designs in pharmacotherapy. Pharmacoepidemiol Drug Saf 2001;10:259–62.
55. **Whittemore AS, McGuire V**. Observational studies and randomized trials of hormone replacement therapy: what can we learn from them? Epidemiology 2003;14:8–10.
56. **Hennekens CH, Buring JE**. Statistical association and cause-effect relationships. In: Mayrent SL, ed. Epidemiology in

medicine. Philadelphia, PA: Lippincott Williams & Wilkins, 1987:30–53.

57. **Rothamn KJ, Greenland S, Poole C, Lash TL, eds**. Causation and causal inference. In: Modern epidemiology. Philadelphia,

PA: Lippincott Williams & Wilkins, 2008:5–31.

58. **Bhopal R, ed**. Cause and effect: the epidemiological approach. In: Concepts of epidemiology, 2nd ed. New York, NY: Oxford University Press, 2008:123–62.

Appendix 1. Brief List of Resources Detailing Good Survey Research Practices

Textbooks

    Aday LA, Cornelius LJ. Designing and conducting health surveys: a comprehensive guide, 3rd ed. San Francisco, CA: Jossey-Bass, 2006.

    Dillman DA, Smyth JD, Christian LM. Internet, mail, and mixed-mode surveys: the tailored design method, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2009.

    Fowler FJ. Survey research methods, 3rd ed. Thousand Oaks, CA: Sage Publications, 2002.

Journal articles

    Draugalis JR, Coons CJ, Plaza CM. Best practices for survey research reports: a synopsis for authors and reviewers. Am J Pharm Educ 2008;72:article 11.

    Kelley K, Clark B, Brown V, Sitzia J. Good practices in the conduct and reporting of survey research. Int J Qual Health Care 2003;15:261–6.

    Rubenfeld GD. Surveys: an introduction. Respir Care 2004;49:1181–5.