CHAPTER 4

NONEXPERIMENTAL QUANTITATIVE RESEARCH

GABRIELLA BELLI

KEY IDEAS

- The distinction between experimental and nonexperimental research rests on the manipulation of treatments and on random assignment.
- Any quantitative study without manipulation of treatments or random assignment is a nonexperimental study.
- Nonexperimental research is used when variables of interest cannot be manipulated because they are naturally existing attributes or when random assignment of individuals to a given treatment condition would be unethical.
- Numbers are used to represent different amounts of quantitative variables and different classifications of categorical variables.
- Nonexperimental studies may be classified along two dimensions: one based on the purpose of the study and the other on the time frame of the data collection.
- Evidence of a relationship is not convincing evidence of causality.
- Alternative explanations for results in nonexperimental research should be explored and ruled out.

NOTE: My thanks to Professor Bill Frakes, from the Computer Science Department at Virginia Tech, and to students, including many from my Research Methods class in Fall 2007, for reviewing a prior draft of this chapter. Their insightful comments and suggestions helped improve this version. I take responsibility for any remaining elements of confusion that may remain.

OVERVIEW OF NONEXPERIMENTAL RESEARCH

QUANTITATIVE RESEARCH is empirical, using numeric and quantifiable data. Conclusions are based on experimentation and on objective and systematic observations. Quantitative research may be divided into two general categories: experimental and nonexperimental. The essential elements of experimental research, which was discussed in detail in the previous chapter, are presented here first as a contrast to nonexperimental research. A primary goal for experimental research is to provide strong evidence for cause-and-effect relationships. This is done by demonstrating that manipulations of at least one variable, called the treatment or independent variable (IV), produce different outcomes in another variable, called the dependent variable (DV). An experimental study involves at least one IV that is manipulated or controlled by the researcher, random assignment to different treatment conditions, and the measurement of some DV after treatments are applied. Any resulting differences in the DV across the treatment groups can then be attributed to the differences in the treatment conditions that were applied.

In contrast to experimental research, nonexperimental research involves variables that are not manipulated by the researcher and instead are studied as they exist. One reason for using nonexperimental research is that many variables of interest in social science cannot be manipulated because they are attribute variables, such as gender, socioeconomic status, learning style, or any other personal characteristic or trait. For example, a researcher cannot randomly place individuals into different groups based on gender or learning style because these are naturally existing attributes.

Another reason to use nonexperimental research is that, in some cases, it would be unethical to randomly assign individuals to different treatment conditions. A classic example of this is that one could not study the effects of smoking by randomly assigning individuals to either a smoking or a nonsmoking group for a given number of years. The only ethical way to investigate the potential effects of smoking would be to identify a group of smokers and a group of nonsmokers and compare them for differences in their current state of health. The researcher, however, would also need to take other variables into account, such as how long people had smoked, their gender, age, and general health level. To do so would be important because the researcher cannot take for granted that the groups are comparable in aspects other than smoking behavior. This is in contrast to experimental groups, which, due to the process of random assignment, start out equal in all respects except for the treatment condition in which they are placed. In nonexperimental research, groups based on different traits or on self-selection, such as being or not being a smoker, may differ for any number of reasons other than the variable under investigation. Therefore, in nonexperimental studies, one cannot be as certain as in experimental studies that outcome differences are due to the independent variable under investigation. The researcher needs to consider possible alternative explanations, to jointly analyze several variables, and to present conclusions without making definitive causal statements.

In this chapter, you will learn how to characterize nonexperimental studies that do not rely on either manipulation of variables or random assignment of subjects to groups. Different types of nonexperimental studies will be explained, and you will learn how to characterize them using a two-dimensional classification system. By the end of the chapter, you will understand the basic elements of nonexperimental studies, as well as the rationale for their use. Nonexperimental research examples, including published studies, will be incorporated into the discussion to facilitate understanding. At the end of the chapter, text and Web resources are provided to help you locate supplemental materials and additional information.

VARIABLES AND THEIR MEASUREMENT

To facilitate reading the remainder of the chapter, a brief review of variables and some of their different aspects is presented. A **variable** is any characteristic or attribute that can differ across people or things; it can take on different values. Some variables are inherent traits, such as gender or height. Others may vary due to experimenter manipulation, such as treatment groups of drug versus placebo, or due to self-selection, such as attending a two- or a four-year college. In quantitative research, variables are measured in some way and those numerical values are then used in statistical analyses. The nature of variables is important because, to some extent, it dictates the way research questions are asked and which analysis is used.

One basic distinction is that variables can be either categorical or quantitative. **Categorical variables** are those that differ across two or more distinct categories. The researcher assigns arbitrary numbers to the categories, but the numbers have no interpretable numerical meaning. For example, for categories of the variable "employment status," we could assign the value "1" to employed full-time, "2" to employed part-time, and "3" to not employed. Additional examples of categorical variables that are individual traits are gender, ethnicity, and learning style; some that are self-selected are marital status, political party affiliation, and field of study.

Quantitative variables can be measured across a scale, their numeric values have meaning, and they can be subjected to arithmetic operations. The following are all examples of quantitative variables: age, height, weight, grade point average (GPA), job satisfaction, and motivation. There is an important distinction between the first three and the last three variables in this list. For such variables as age, height, and weight, zero is a meaningful value that indicates the absence of the characteristic being measured, as in something that is brand new or has no weight. The numbers have interpretable meaning. We know what five years or five feet means because there is no arbitrariness about these values or how to interpret them.

In contrast, zero is an arbitrary value for variables such as GPA, satisfaction, or motivation. A zero motivation score does not mean one has no motivation, but merely that one attained the lowest possible score for the particular instrument being used. GPA in most schools in the United States is given on a continuum from 0.0 to 4.0 but, for example, at the Massachusetts Institute of Technology (MIT), it goes from 0.0 to 5.0 (see GPA calculation and unit conversion in MIT Web page at http://web.mit.edu/registrar/gpacalc.html). The International Baccalaureate grades range from 1 to 7, based on a rubric developed from the standardized curriculum.

For another example, consider measurements for temperature. The freezing point of water is represented as zero on a Celsius thermometer, but as 32 on a Fahrenheit thermometer. In neither case does a zero represent the absence of temperature. In each case, we understand what the numbers mean because specific interpretations have been assigned to them.

Interpretation of different grading schemes or thermometers is possible because of commonly understood unit descriptors. This is not so for such variables as job satisfaction or motivation, where scores are arbitrary and depend on the measurement instrument being used and how it has been designed. Typically, such scores are the sum or the average of responses to a set of items. The items may be statements, constructed so that all are related to the variable to be measured, and responses are often, but not always, on a Likert scale from 1 (strongly agree) to 5 (strongly disagree). The terms **scale** and **index** are often used to describe such sets of related items that, together, produce a score about some characteristic or phenomenon. For example, the Multidimensional Job Satisfaction Scale (Shouksmith, Pajo, & Jepsen, 1990) contains eleven different subscales, each a multi-item scale measure of a different dimension of job satisfaction. Another instrument, the Job Satisfaction Survey (Spector, 1985), consists of nine four-item subscales to assess employee attitudes about the job. As you can see from this example, different researchers developed different measures of the same construct, job satisfaction.

Exact interpretation of a scale score's value, or measure, for variables such as motivation or satisfaction is not important. What is important is to know that the higher the score, the more one has of the characteristic being measured and vice versa. One could, for example, examine whether males or females had higher levels of job satisfaction or if people with higher levels of job satisfaction also tended to have higher levels of motivation. To be confident of results, it is also important to know that the measures being used are reliable and have been validated.

Reliability relates to the consistency or dependability of a measure. Basically, if it is reliable, you can be confident that all the items that make up the measure are consistent with each other and that, if you were to use the measure again with the same individuals, they would be rated similarly to the first time. **Validity** relates to whether it is measuring what we intend it to measure, and represents the overarching quality of the measure. The purpose of using the measure is an important consideration in evaluating validity because it could be valid for one use but not for another. These concepts are complex and beyond the scope of this chapter (see Trochim, 2005 for a very understandable description of validity and reliability of measures). As a consumer of research, you should at least be aware of them and look for how research authors deal with these concepts. Do they describe their measures in detail and provide some indication of reliability and validity?

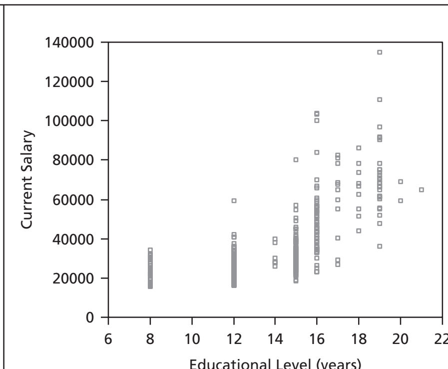
Defining Variables

Although some variables are inherently categorical or quantitative, others may be defined in either way. Imagine, for example, that you are interested in measuring the education level of a group of individuals. You could do this categorically, by defining education as "highest degree earned" and using five values representing none, high school, college, masters, or doctorate as different levels of education. Or, you could do this quantitatively by defining education as "number of years of schooling," where the resulting values would be meaningfully interpreted. This distinction is important if one is interested in studying the relationship between educational level and salary, a quantitative variable, because it relates to how the data might be analyzed and how research questions would be phrased. Using the categorical definition, you could compare the median salary value across the five categories of "highest degree earned." The **median** represents the midpoint when all the salaries are listed from lowest to highest. One could then determine if there were any appreciable differences in salary across the five groups and whether more education (represented by having a higher degree) corresponded to higher salary.

Using the quantitative definition, you could graph the two variables in a **scatter plot** or compute a **correlation coefficient** (a measure of strength and direction of relationship for two variables) for the number of years of schooling and salary. The first would provide a visual representation of their relationship and the second a numerical one. Figure 4.1 shows how resulting data might be depicted in the two cases described. The table shows the number of people in each group and their median salary. The scatter plot shows all the data points. The correlation for this data set is 0.66. Correlation

FIGURE 4.1. ***Two Representations of the Relationship Between Salary and Education Level***

Highest Degree	N	Median Salary
Doctorate	30	68,438
Master's	20	65,938
Bachelor	181	33,150
High School	190	24,975
None	53	24,000
Total	474	28,875



Education is measured as a categorical variable (highest degree). The size of each group (N) and the median salary are given in the table.

Education is measured as a quantitative variable (number of years in school). Each point in the scatter plot represents years in school and salary for a single individual.

values range from −1 to +1, with zero indicating no relationship and 1 indicating either a negative or a positive perfect relationship depending on the sign. We could say these data showed a moderate positive relationship. Fewer years of schooling tend to correspond to lower salaries and more schooling to higher salaries.

Phrasing Questions

In the first case demonstrated in Figure 4.1, you would be comparing groups with different levels of education on some measure (salary), and in the second case, you would be relating two sets of numeric scores (years and salary). The research questions of interest in the two cases would be: (1) how do groups, based on highest degree earned, differ from each other with respect to salary? and (2) how does number of years of schooling relate to salary? Phrased generically, the key questions in the two situations are: How do groups differ from each other on some measure? How are the variables related to each other? The distinction between these two cases depends only on the fact that education was conceptualized as either categorical or quantitative and not on the nature of the relationship involved.

REFLECTION QUESTIONS

By now, you should be able to:

1. Describe the difference between experimental and nonexperimental studies
2. Give an example of an independent and a dependent variable within the context of a research question
3. Give an example of a categorical and a measured, quantitative variable

CLASSIFYING NONEXPERIMENTAL RESEARCH

In the literature on experimental studies, there is agreement on the distinction between true- and quasi-experiments. Although both involve treatment manipulation, **true-experiments** use random assignment of subjects to groups *and* random assignment of groups to treatments. **Quasi-experiments** use preexisting intact groups, which are randomly assigned to treatment conditions.

For nonexperimental designs, there appears to be no consistent agreement on typology. In 1991, Elazar Pedhazur and Liora Schmelkin stated that "there is no consensus regarding the term used to refer to designs" which were presented in their chapter on nonexperimental designs (p. 305). Two commonly used terms for nonexperimental studies are "correlational research" and "survey research." However, the term *correlation* relates more to an analysis strategy than to a research design and the term *survey* describes a method of gathering data that can be used in different types of research.

Ten years later, Burke Johnson (2001) came to the same conclusion. Based on a review of twenty-three leading methods textbooks in education and related fields (thirteen explicitly from education and the rest from anthropology, psychology, political science, and sociology), he found little consistency in how nonexperimental studies were classified. He discovered over two dozen different labels being used, sometimes with slight variations in the wording. The most frequently used labels in these texts were survey (twelve times), correlational (ten times), descriptive (eight times), and causal-comparative (five times). The result of my informal review of six additional research methods texts was consistent with Johnson's findings.

In an attempt to remedy this confusion, Johnson (2001) proposed a categorization scheme consisting of two basic dimensions, each with three categories. The first dimension represents a characterization of the basic goal or main purpose for conducting the nonexperimental quantitative study. The second dimension allows the research to be classified according to the time frame in which data were collected. These two dimensions will be presented here and discussed separately in the next two sections. In your reading of published articles or research methods textbooks, you will probably encounter other terms for nonexperimental research. You may want to read Johnson (2001) to familiarize yourself with these terms and with the problems that arise because of their use.

Classification Based on Purpose (Dimension 1)

The categories of the first dimension for classifying nonexperimental studies, which are based on the main purpose of the study, are:

1. **Descriptive** nonexperimental research, in which the primary focus for the research is to describe some phenomenon or to document its characteristics. Such studies are needed in order to document the *status quo* or do a needs assessment in a given area of interest.
2. **Predictive** nonexperimental research, in which the primary focus for the research is to predict some variable of interest (typically called the **criterion**) using information from other variables (called **predictors**). The development of the proper set of predictors for a given variable is often the focus of such studies.
3. **Explanatory** nonexperimental research, in which the primary focus for the research is to explain how some phenomenon works or why it operates. The objective is often to test a theory about the phenomenon. Hypotheses derived from a given theoretical orientation are tested in attempts to validate the theory.

The three categories could be seen as answers to the question: Was the main purpose of the research to describe a phenomenon, to study how to predict some future event, or to understand how something operates or what drives it?

To help explain these three categories, consider the use of exit interviews. Such interviews are often conducted by organizations with employees who leave or by school systems with departing teachers and graduating seniors. An exit interview study can be **descriptive** if the purpose is to collect data in order to get a comprehensive picture of reasons for employees leaving their organization or school. These descriptions might be used to determine if people leave for reasons related to the organization or for personal reasons. On the other hand, the study would be **predictive** if exit data were collected and then related to hiring data for the same individuals for the purpose of using the results to screen potential employees and hiring people who might be less likely to leave. Finally, the study would be **explanatory** if the data were analyzed with the purpose of testing hypotheses about how personal characteristics might be related to employee or student feelings about their organization or school.

A good example of a published **descriptive** study is the *39th Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes toward the Public Schools* (Rose & Gallup, 2007). Begun as an effort to inform educators, the annual survey now provides information that has policy implications. Although the accumulated database can be used to track changes in attitudes about Pre-K–12 schooling over a long period of time, the design for each yearly survey is purely descriptive in terms of its purpose. Results are a descriptive representation of how the general public feels about different aspects of public schools.

A study by Leslie Halpern and Thomas Dodson (2006) to develop a set of indicators that could identify women likely to report injuries related to intimate partner violence is an example of a **predictive** study. They tried to develop markers that could be used in hospital settings to make predictions about likelihood of intimate partner violence. They identified two variables as potential predictors: injury location and responses to a standard screening questionnaire. They included them, along with demographic variables, in developing a prediction model.

An **explanatory** study was done to examine the relationships among the variables of attachment, work satisfaction, marital satisfaction, parental satisfaction, and life satisfaction (Perrone, Webb, & Jackson, 2007). This research was informed by attachment theory, which describes "parental attachment as a stable connection that provides a feeling of safety and security for the child" (p. 238). The researchers used five published instruments and present a very good description of reliability and validity for each one.

Classification Based on Time (Dimension 2)

The categories of the second dimension for classifying nonexperimental research, which refer to time, are:

1. **Cross-sectional** research, in which data are collected at one point in time, often in order to make comparisons across different types of respondents or participants.
2. **Prospective** or **longitudinal** research, in which data are collected on multiple occasions starting with the present and going into the future for comparisons across time. Data are sometimes collected on different groups over time in order to determine subsequent differences on some other variable.
3. **Retrospective** research, in which the researcher looks back in time using existing or available data to explain or explore an existing occurrence. This backwards examination may be an attempt to find potential explanations for current group differences.

These categories could be seen as answers to the question: Were the data collected at a single time point, across some time span into the future, or were already existing data explored? You could think of them as representing the past (retrospective), present (cross-sectional), and future (prospective) with respect to timing of data collection. As an example, suppose you were interested in assessing differences in college students' attitudes toward potential careers. In a **cross-sectional study**, you might take a random sample of first-year college students (freshmen) and fourth-year college students (seniors) and compare their attitudes. Your purpose might be to show that more mature students (seniors) view career options differently from less mature students (freshmen).

Now consider assessing career attitudes in a **prospective study**. There are actually three options: trend, cohort, or panel study. To distinguish among these three approaches, think of a four-year prospective study starting in 2008 with college freshmen. The population of interest is all college freshmen in the United States. In 2008, a **random sample** of college freshman is taken for all three approaches. Table 4.1 describes the samples in the subsequent three years for each approach. In the trend study, the same *general population* (college freshmen) is tracked. In the cohort study, the same *specific population* (college freshmen in 2008) is tracked. In the panel study, the *same individuals* are tracked. One of the advantages of a panel study is that you can look for changes and not simply report on trends. A disadvantage is that you have to start with a fairly large sample due to attrition over time, particularly for a lengthy study.

An example of a **retrospective** study could be an examination of the educational background and experience of very successful teachers and less successful teachers. The idea is to look backward in time and examine what differences existed that might provide an explanation for the present differences in success. To the extent that such a study needed to depend on people's memories of relevant background information, it would be less accurate than if prior data were available for examination.

For a published example, consider one question addressed by Michael Heise (2004), which was whether key actors in a criminal court case view case complexity in the same way. The results of his **cross-sectional** comparison of three key actor groups (juries, attorneys, and judges) suggest that they do possess slightly different views on whether crimes are complex.

Examples of both prospective and retrospective research are based on the *Nurses' Health Study*, a large scale longitudinal study started in 1976 with a mailed survey of 121,700 female registered nurses between thirty and fifty-five years of age who lived in eleven states. Descriptive information about risk factors for major chronic diseases and related issues were gathered every two years. Although most of the information gathered

TABLE 4.1. Description of Samples After Initial 2008 Sampling of College Freshmen

	2009	2010	2011
Trend	New sample—college freshmen	New sample—college freshmen	New sample—college freshmen
Cohort	New sample—college sophomores	New sample—college juniors	New sample—college seniors
Panel	Same sample from 2008, who are now sophomores	Same sample from 2008, who are now juniors	Same sample from 2008, who are now seniors

was identical, new questions were added periodically. The *Nurses' Health Study* Web page (www.channing.harvard.edu/nhs) contains a complete list of publications based on these data.

One such study was conducted by Francine Laden et al. (2000). They examined the responses from the 87,497 women who answered newly included questions about lifetime use of electric blankets and heated waterbeds. Using data from the larger study, Laden and her colleagues focused their attention on the relationship between electric blanket use and breast cancer from both a prospective and retrospective view. This was done because electric blanket use is a source of electric and magnetic fields (EMFs) exposure, and EMF exposure had been hypothesized to increase the risk of breast cancer. The relevant year is 1992, when information about use of electric blankets and waterbeds was first documented. For the **prospective** part of their study, they considered women who had not been diagnosed with cancer as of 1992 and analyzed the occurrence of breast cancer from 1992 to 1996 for groups according to electric blanket or waterbed usage. For the **retrospective** part, they used records from 1976 to 1992, considering only women who were cancer free in 1976. In the prospective part of the study "exposure to electric blankets and waterbed use was assessed prior to the occurrence of breast cancer," while in the retrospective analysis "exposure was ascertained after diagnosis" (Laden et al., 2000, p. 42).

Retrospective studies may be based on past records, as in the previous example, or on retrospective questions, that is, on questions about past behaviors or experiences. Merely using already existing data, however, does not make it retrospective. The key distinction is the study's purpose. Are you looking backwards to discover some potential cause or explanation for a current situation, or are you using data from one point in time to predict data from a later time? Notice that Laden and her colleagues (2000) used preexisting data for both retrospective and prospective studies. For the prospective part, women who had not been diagnosed with cancer in 1992 were divided into groups based on whether they did or did not use electric blankets, and the groups were then compared with respect to breast cancer incidents by 1996. For the retrospective part, they divided the women into two groups based on whether they had or had not been diagnosed with cancer as of 1992 and then compared them in terms of reported prior use of electric blankets.

Combining Classification Dimensions

When used together, Johnson's two dimensions (2001) combine to form a 3×3 design for a total of nine distinct categories that may be used to describe nonexperimental research. Examples of all nine may be found in the *National Education Longitudinal Study of 1988* (NELS:88), which was a large-scale data collection effort. A nationally representative sample of eighth graders were first surveyed in 1988, with subsequent follow-up surveys every two years until 1994, and then once again in 2000. The National Center for Education Statistics' Web page (http://nces.ed.gov/surveys/nels88) describes this study, and also provides an annotated bibliography of research done using the various data sets. Depending on which data were selected for each study and the study

TABLE 4.2. Articles Classified According to Both Research Objective and Time of Dimensions

	Retrospective	Cross-Sectional	Prospective
Descriptive	*Type 1*	*Type 2*	*Type 3*
	Behavioral responses of substance-exposed newborns: a retrospective study (Higley & Morin, 2004)	Criminal case complexity: An empirical perspective (Heise, 2004)	The stability of undergraduate students' cognitive test anxiety levels (Cassady, 2001)
Predictive	*Type 4*	*Type 5*	*Type 6*
	Electric blanket use and breast cancer in the nurses' health study (Laden et al., 2000)	A predictive model to identify women with injuries related to intimate partner violence (Halpern & Dodson, 2006)	Electric blanket use and breast cancer in the nurses' health study (Laden et al., 2000)
Explanatory	*Type 7*	*Type 8*	*Type 9*
	A further look at youth intellectual giftedness and its correlates: values, interests, performance, and behavior (Roznowski, Reith, & Hong, 2000)	Relationships between parental attachment, work and family roles, and life satisfaction (Perrone, Webb, & Jackson, 2007)	Thirty-year stability and predictive validity of vocational interests. (Rottinghaus, Coon, Gaffey, & Zytowski, 2007)

purpose, different NELS:88 studies might be classified using all nine of the purpose by time frame classifications. To help clarify this cross-classification scheme, Table 4.2 gives the titles of articles representing each type, which are then described.

Type 1—Descriptive retrospective. Using retrospective chart review, Anne Marie Higley and Karen Morin (2004) described the behavior of infants whose mothers had a drug history. Their findings supported the use of an assessment tool to guide parents in providing a supportive care environment to help infants recover.

Type 2—Descriptive cross-sectional. This study was discussed earlier as an example of a cross-sectional study. It is descriptive because the goal was to document the extent to which juries, attorneys, and judges held similar or different views about a case. The results have implications for legal reform efforts.

Type 3—Descriptive prospective. This was an investigation of the stability of test anxiety measures over time and testing formats, with data collected at three time points in an academic semester, therefore making it prospective. The purpose for the description was to determine if test anxiety was a stable condition or if it is necessary to include a test anxiety measure with every test in a longitudinal study. Results indicated that it is not necessary to measure anxiety with every test; it is only necessary to measure anxiety in one test-taking situation.

Type 4—Predictive retrospective and **Type 6—Predictive prospective**. The two parts of this study were described earlier as examples of retrospective and prospective studies. Both parts were predictive in nature, using a backward and a forward perspective to determine the extent to which electric blanket and waterbed use could be used to predict breast cancer. Although results did not exclude small risks, neither analysis supported an association between breast cancer risk and use of electric blankets and waterbeds.

Type 5—Predictive cross-sectional. In this study, discussed as an example of a predictive study, a one-time data collection was used. The authors' aim was to develop and validate a predictive model. They subdivided their sample, using one group to develop their model and the second group to validate, or test it. Their work produced a predictive and validated model of three components: risk of self-report of intimate partner violence related injury, age, and race. The researchers then hypothesized that these three variables could be used to develop a protocol to assist in the early diagnosis of intimate partner violence in an emergency department and outpatient clinical setting.

Type 7—Explanatory retrospective. This study was explanatory because a goal was to further previous work on giftedness and knowledge and understanding of several related variables. The data came from the High School and Beyond database, a longitudinal study with baseline information on 14,825 students who were high school sophomores in 1980. The data for this study included the base year and the third follow-up survey, four years later, after graduation. The data set "allowed for more comparisons than could reasonably be included in a single study. Variables were chosen that would either serve to replicate previous findings or expand psychological and behavioral profiles of gifted male and female students into more detail" (Roznowski, Reith, & Hong, 2000, p. 96). A retrospective conclusion was that educational attainment differences of gifted males and females had their origins in the early high school years.

Type 8—Explanatory cross-sectional. Already discussed as an example of an explanatory study, this study was based on data from the fifteenth annual survey of a longitudinal study that started in 1988 with 1,724 participants. About 1,200 participants were lost in the first three years. Only 108 participants were left for this study, which shows the dramatic attrition that can happen in a longitudinal study. Although the data were from a longitudinal study, these authors only used the fifteenth year's data, thereby making it cross-sectional.

Type 9—Explanatory prospective. The authors suggested that "Assessing the predictive validity of an interest inventory is essentially answering the question, 'Do early interest scores match one's future occupation?'" (Rottinghaus et al., 2007, p. 7). To answer this question, they did a thirty-year follow-up of 107 former high school juniors and seniors whose interests were assessed in 1975. The first author had collected the initial data. Their results extend research on vocational interests, indicating that interests were fairly stable even after such a long time span.

REFLECTION QUESTIONS

1. How do descriptive, predictive, and explanatory studies differ?
2. How do retrospective, cross-sectional, and prospective studies differ?
3. Find several recent articles in your field of study where a nonexperimental design was used. Classify their main purpose as being descriptive, predictive, or explanatory and classify the time dimension as retrospective, cross-sectional, or prospective.

CAUSAL EXPLANATIONS AND NONEXPERIMENTAL STUDIES

Using Johnson's classification system (2001), many nonexperimental studies are either descriptive or predictive. For those, the notion of causation is not relevant. However, a goal for many explanatory nonexperimental research studies is to explore potentially causal relationships. A causal relationship is one in which a given action is likely to produce a particular result.

The terms *independent* and *dependent* refer to the different roles variables play in experimental studies. If a causal relationship exists, then the outcome (the measured DV) depends on, or is a direct result of, the nature of the assigned independent treatment condition. Strictly speaking, these terms are not applicable in nonexperimental research, although they are often used. The more appropriate terms in nonexperimental studies are **criterion** and **predictor** variables, criterion being the presumed outcome of one or more predictor variables. When the intent is to use nonexperimental research to study potential cause-and-effect relationships where experimentation is not possible, the concept of IV and DV may still be of interest, but conclusions about causation that can be made from nonexperimental studies are weaker than those that can be made from true-experimental studies. Additionally, great care needs to be taken to assure that nothing essential has been overlooked.

As explained earlier, the distinction is often made between nonexperimental studies that involve both categorical and quantitative variables and those that involve only quantitative variables. Considering only two variables for the sake of simplicity, an example of the first type of study is a comparison of gender differences in mathematics achievement in high school. Gender, with male and female as the two categories, is considered the independent variable and some mathematics achievement score is the measured dependent variable. Examples where both variables are quantitative might be an examination of the relationship between test scores and time spent studying, or between scores on some measure of motivation and scores on an achievement test. Examples like these, of very simple cases involving only two variables, are neither very interesting nor very informative. Additional variables could be included in order to examine more complex relationships.

No matter which type of design or which type of variable is used, evidence of a relationship would not be convincing evidence of causality. Recall the example described earlier about investigating the relationship between education level and salary and the two ways that education level could be measured. Regardless of whether education level was construed as categorical (highest degree earned) or as quantitative (number of years of schooling), it should not be concluded that one's educational level *caused* or produced a different level of salary. If dramatic differences across the five groups with different degrees were found such that those with higher education had higher median salaries, all that can be concluded is that there was a *relationship* between educational level and salary. This same conclusion would be possible if results indicated a strong positive correlation between years of schooling and salary: that people with fewer years of school tended to have low salaries and people with more years of school tended to have high salaries (see Figure 4.1 for graphical representation of a positive relationship). The scatter plot for a negative relationship would go from the upper left corner to the lower right corner, indicating that low scores on one variable tended to go with high scores on the other variable.

The differences in the wording of the research questions in the previous two cases reflect the nature of the variables used (categorical or quantitative). They would require different analysis strategies, either to test if the median values did differ more than you might expect by chance, or to determine the strength and direction of the relationship. Differences in wording or analysis do not, however, reflect any difference in the nature of the relationship between the variables. Explanatory nonexperimental research articles often have conclusions phrased in causal language. Therefore, the next section is a review of the essential elements needed to establish cause-and-effect relationships and a discussion of their applicability to nonexperimental studies.

Requirements for Causality

There are three conditions necessary in order to be able to argue that some variable X (the presumed independent) causes another variable Y (the presumed dependent).

1. The two variables X and Y must be related. If they are not related, it is impossible for one to cause the other. For nonexperimental research, that means that it must be demonstrated that differences in X are associated with differences in Y.
2. Changes in X must happen before observed changes in Y. This is always the case when X is a manipulated treatment variable in an experiment. But establishing that a cause happened before an effect needs to be documented in some way or logically explained in nonexperimental studies. This is impossible to do when the data are cross-sectional and collected simultaneously.
3. There is no possible alternative explanation for the relationship between X and Y. That is, there is no plausible third variable that might explain the observed relationship between X and Y, possibly having caused both of them.

In nonexperimental studies, the first requirement can be established easily with correlational analyses. The second could also be established if longitudinal data are used so that predictor variables are measured before the criterion. The third requirement is more difficult to demonstrate. To do so requires a thorough knowledge of the literature and the underlying theory or theories governing the topic being investigated, logical arguments, plus testing and ruling out of alternative possibilities.

The fact that two variables are related does not inform us of which one influences the other. There are at least three reasons why two variables could be related and it is not possible to know from the correlation which one is the correct reality. Three potential explanations are: (1) that X causes or influences Y, (2) that Y causes or influences X, or (3) that Z, a third variable, causes both X and Y. Consider the following headline: "Migraines plague the poor more than the rich." It could be argued that the stresses of living in poverty and other poverty-related conditions could trigger migraine headaches. It could also be argued that migraines cause one to miss work and eventually lose employment, thereby inducing poverty for a subset of individuals prone to migraines. Which is the correct interpretation? It is impossible to tell.

Although there is no formal way to prove causation in nonexperimental research, it may be possible to suggest it. This is done through careful consideration, by referring to the three conditions for cause, by presenting logical arguments, and by testing likely alternatives in order to make a case for the *likely* conclusion of a causal relationship. One must be careful, however, not to phrase conclusions as proof of causation.

Ruling Out Alternative Hypotheses

To demonstrate the process for ruling out alternative hypotheses, we will use a medical example. Consider the process a doctor goes through in diagnosing a new patient's illness. First, the doctor considers the symptoms. The list of symptoms is used to select potential problems with similar symptoms and to rule out problems with different symptoms. Tests are ordered to confirm the most likely diagnosis and remedies are tried. If the test results are negative or the remedies do not work, then the original diagnosis is discarded, and other possible diagnoses are considered and tested. How does this process relate to research? The first step is matching observations (the reported symptoms) to theory (known symptoms for an illness). The second step is to test a hunch or tentative hypothesis (initial diagnosis) and rule out alternative hypotheses (other potential diagnoses). The process continues until a reasonable conclusion is reached. The analogy breaks down because, ideally, the correct diagnosis is made and the patient is cured, although results are never as conclusive in nonexperimental studies.

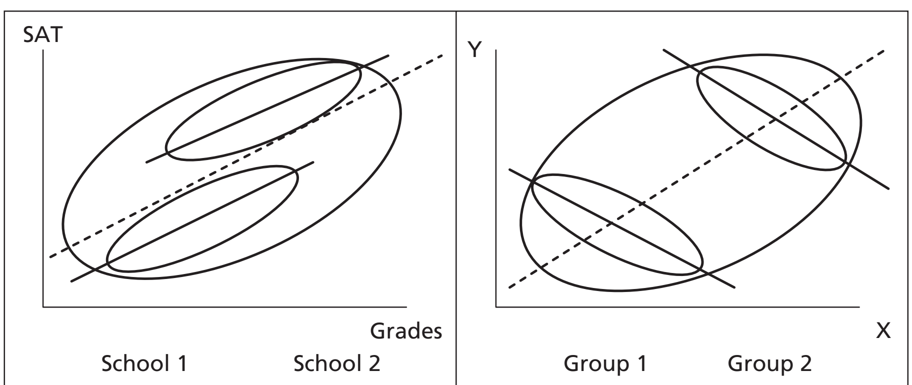
Given a theory that is driving the research, how does one rule out potential alternative hypotheses? One way is to consider all likely **confounding** or **lurking variables**. In an experimental study, two variables are confounded when their effects on a dependent variable cannot be distinguished. The following example, although purely correlational, should clarify the concept of confounding or lurking variables.

One would expect that grades and standardized tests, such as SAT scores, would be related more to each other than they would to socioeconomic status (SES). In many studies, however, SES and SAT appear to have a much stronger relationship than do grades and SAT. Rebecca Zwick and Jennifer Green (2007) explored reasons for such results with data from a random sample of 98,391 students from 7,330 high schools. They performed two different analyses. In the first analysis, they found the correlation for grades and SAT for the entire sample and, in the second analysis, they did so for each school individually and then averaged the school-level results to get one overall measure of relationship. The second analysis produced a much stronger relationship between grades and SAT scores than did the first analysis. This is because the first analysis ignored the fact that there are school-level differences in SES as well as other variables.

Figure 4.2 should help you visualize this discussion. In part A, the two smaller ovals represent a scatter plot of scores for two schools, where both grades and SAT scores tend to be higher in School 2 than in School 1. The lines bisecting these two ovals provide a linear representation of the relationship between the variables within each school and are called **regression lines**. Both ovals are rather narrow in width, being fairly close to their regression lines, and thereby give a visual representation of a relatively strong positive relationship between grades and SAT *within* each school. The larger oval represents the relationship between grades and SAT scores as it would appear across or *between* schools, that is, if school membership were ignored in the analysis. It is much more spread out around its regression line (the dotted line), erroneously indicating a much weaker relationship between grades and SAT. The two smaller ovals correspond to Zwick and Green's second analysis (2007) and the larger oval to their first analysis. Ignoring the differences between the schools confounds the relationship between grades and SAT being investigated.

Part B of Figure 4.2 shows a worst-case scenario of ignoring a lurking variable. Suppose the relationship between two variables, X and Y, is negative for each of two groups. This is shown by the two smaller ovals, where lower scores on X tend to go with higher scores on Y and vice versa within each group. Ignoring groups,

FIGURE 4.2. ***Representation of Effects of Confounding Variables***



A. Fairly strong positive relationship between grades and SAT within each school. Weak relationship when school membership is ignored.

B. Fairly strong negative relationship between X and Y within each group. A seemingly positive relationship when group membership is ignored.

however, would produce a positive relationship, which would be a completely wrong conclusion.

REFLECTION QUESTIONS

By now, you should be able to

1. List and explain three essential requirements to argue cause
2. Explain why even a strong correlation does not imply causation
3. Describe why ruling out alternative hypotheses is important.
4. Find one or two nonexperimental studies in your field of study where hypotheses were tested or where a theory was explored. What extraneous variables or potential alternative hypotheses were discussed? Can you think of others that were not discussed? How might inclusion of those variables have changed results?

ANALYSIS AND INTERPRETATION IN NONEXPERIMENTAL STUDIES

Data analyses in nonexperimental studies depend on both the goal for the study and the nature of the variables in the data set. Almost any analysis may be possible and a useful presentation is not reasonable here. There are ample books and sources for details about statistical methods and their use. A few examples are given at the end of the chapter; also see the discussion on understanding quantitative data in Chapter Six.

You need to be aware of the basic distinction between descriptive and inferential statistics. **Descriptive statistics** involve summarizing and describing quantitative information in meaningful ways. For example, a mean, or arithmetic average, is a statistic used to describe a central value for a set of numbers. **Inferential statistics** are used to make conclusions beyond the data collected and to test hypotheses. Statistical tests are used to make conclusions about populations based on results from random samples or to determine the probability that results are not due to random chance.

Interpretation of results in nonexperimental studies should be consistent with the nature of the work, which is based on nonmanipulated variables. Therefore, conclusions about cause and effect are not appropriate in *any* nonexperimental study. As you read empirical articles, you should be attuned to how conclusions are discussed and be wary of causal language. Robinson, Levin, Thomas, Pituch, and Vaughn (2007) reviewed 274 empirical articles in five teaching-and-learning research journals in 1994 and 2004. They recorded causal and noncausal language use in abstracts and discussion sections. Their two main conclusions were: (1) experimental articles in teaching-and-learning declined in the ten-year span, and (2) on average, the use of causal conclusions made in nonexperimental and qualitative studies increased. They conclude by saying that "as journal readers, we have an obligation to search an article for information about how the data were collected so we are not unduly influenced by unwarranted conclusions" (Robinson et al., 2007, p. 412). Ideally, after studying this chapter you will be able to search through articles for information about how the study was conducted and use that to consider conclusions.

SUMMARY

The goal for this chapter was to present adequate information about nonexperimental designs so that a practitioner could read the literature and have a basic understanding of methods used. Nonexperimental research is described in many ways and covers any quantitative study that does not have manipulated variables or random assignment. A topic of research interest can be modified to serve alternative purposes, and data can be collected over different time frames. The two-dimensional classification system presented here should help you categorize articles. Reading any of the articles listed in Table 4.2 that are of interest to you could be useful in understanding why it was classified according to the two dimensions given. A good place to start, with a relatively straightforward example, would be the Cassady (2001) article, which is an example of Type 3, a descriptive prospective study. A good exercise would be to find other nonexperimental studies and classify them according to the two dimensions of purpose and time of data collection.

A key to understanding published research is to identify the goal of the research, evaluate what was done in relation to that goal, and consider aspects and variables that may have been overlooked. Most important, consider the language used in published works and be skeptical if overzealous researchers present their nonexperimental results in causal terms. Regardless of what type of research is presented, be a wary consumer.

KEY TERMS

attribute variables
categorical variables
confounding or lurking variables
correlation coefficient
criterion
cross-sectional research
dependent variable
descriptive nonexperimental research
descriptive statistics
experimental research
explanatory nonexperimental research
independent variable
index
inferential statistics
median
nonexperimental research
predictive nonexperimental research
predictors
prospective or longitudinal research
quantitative variables
quasi-experiments
random assignment
random sample
regression line
reliability
retrospective research
scale
scatter plot
true-experiments
validity
variable

FURTHER READINGS AND RESOURCES

Suggested Readings

Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA: Pine Forge Press.

This basic text, discussing an analysis technique often used in nonexperimental studies, is written in an understandable manner, using examples from social science research literature to develop the concepts.

Johnson, R. B., & Christensen, L. See lecture in Chapter Eleven: Nonexperimental quantitative research, based on *Educational Research: Quantitative, Qualitative, and Mixed Applications*. Retrieved March 13, 2008, from www.southalabama.edu/coe/bset/johnson/2lectures.htm.

Discusses steps in nonexperimental research, ways to control extraneous variables in nonexperimental research, and Johnson's classification scheme for nonexperimental research, and provides a graphic description of controlling for a third variable.

Locke, L. F., Silverman, S. J., & Waneen, W. S. (2004). *Reading and understanding research* (2nd ed). Thousand Oaks: Sage.

Although this book deals with research in general, it is an easily understandable resource with good examples to help you read and understand published research articles. Aimed at consumers of research, the approach is nontechnical and user-friendly.

Lowry, R. (1999–2008). *Concepts and applications of inferential statistics*. Retrieved October 10, 2007, from http://faculty.vassar.edu/lowry/webtext.html.

Chapter Three of this free, full-length statistics textbook provides an introduction to linear correlation and regression using examples and diagrams. This is useful for understanding the basic analyses used with nonexperimental data.

Meltzoff, J. (1997). *Critical thinking about research: Psychology and related fields*. Washington, DC: American Psychological Association.

This text should help develop critical thinking skills via research by critiquing exercises of different types of research studies. It combines fundamental content with practice articles.

Trochim, W. M. *The research methods knowledge base* (2nd ed.). Retrieved October 20, 2006, from www.socialresearchmethods.net/kb.

Of particular use is the *Language of Research* part of the *Foundation* section, where types of relationships are clearly described, using simple examples and graphs.