CHAPTER

3 Research Design

The Experimental Model and Its Variations

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

Research design is the plan or blueprint for a study and includes the who, what, where, when, why, and how of an investigation. The research design should flow from the problem formulation and critical issues that were identified for observation. Who is to be investigated—an individual, one group, many groups, organizations, or communities? What is to be investigated—attitudes, behavior, or records? Where is the study to be conducted? Do we wish to look at the past (after-the-fact, post hoc, or a posteriori studies) or the present or to predict the future? Do we want to look at a group once or over time? Why do we wish to do the investigation—to describe, explain, or predict? Finally, how do we design the study so that upon completion we are able to address the hypotheses and present findings that resolve in some manner the research problem?

THE EXPERIMENTAL MODEL

Some people view experiments as involving white-coated scientists with an impressive assortment of equipment tediously studying obscure phenomena in some isolated laboratory. Although such a picture may indeed be accurate in some instances, the experimental model contains many variations and should not be restricted to this stereotypical view. The experimental model will be treated in this chapter as the benchmark for comparison of all other research designs and methods. Most studies of an empirical nature in criminology and criminal justice can be viewed as variations of the experimental model (Weubben, Straits, and Shulman, 1974; Campbell, 1977; and Cook and Campbell, 1979). The research design notation (*X*'s and *O*'s) used in this chapter may at first appear intimidating, but if you give it a chance, you will find it excellent shorthand for dissecting any research study. This chapter contains more "researchese" than any other chapter, but if you bear with it, you will be rewarded by becoming fluent in the language of research.

RESEARCH DESIGN IN A NUTSHELL

In learning something new such as swimming or driving, the initial lessons seem the most difficult; once the foreignness of a new experience is overcome, the rest is relatively easy. Unfortunately, learning research methods is similar. The language of research design in this chapter by its very nature has this same foreignness at first. One philosophy is to teach something by simply doing it, such as throwing the student in the deep end of the pool or on the fast lane on the Capitol Beltway (which, incidentally, is where this writer learned to drive). Rather than following this practice and throwing you in the deep end, let us begin with a short lesson. Read Figure 3.1, even if you do not fully understand it. It provides the "guts" of the entire chapter. As you are reading or after completing the chapter, you may wish to reread it because, if you understand Figure 3.1, you have the underlying logic of research design.

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

X = treatment (independent variable), e.g., Foot Patrol
Y = outcome (dependent variable), e.g., Crime Rate
Z = any rival causal factor (other variables besides X that could really be causing a change in Y), e.g., history, selection bias, testing, etc.
O = observation (some measurement or assessment of dependent variable)
E = equivalence (randomization or matching)
1, 2 = time

Design 1: O_1XO_2 (One Group Before-After Design)

- A precinct with a crime rate of 1,000 (O_1) is exposed to foot patrol (X) for one year and then has a crime rate of 500 (O_2).
- Problem: Other variables (Z) could actually have caused the decline in crime rate (Y) rather than foot patrol (X), e.g., History, Selection Bias, Testing Effects.
- Solution: A better research design (such as Design 2) to control or exclude these rival causal factors (or other variables).

Design 2: EO_1XO_2
EO_1O_2 (Classic Experimental Design)

- Two precincts as similar as possible (E, matching) on relevant characteristics are studied.
  Both are observed (O_1) and have a crime rate of 1,000.
  One precinct receives foot patrol (X) and is the experimental group.
  The other precinct receives no treatment (control group).
  After one year, the experimental precinct has a crime rate of 500 while the control group has a crime rate of 1,000 (or no change). The decrease is attributed to foot patrol.
- Rival causal factors (Z) were controlled for (excluded) by the classic experimental design, for example,
  The change could not have been due to some historical event if we assume that both groups were similar and exposed to the same history.
  The change could not be due to selection bias because we purposely chose two similar or matched precincts (E) and their crime rates were the same from the beginning (O_1).
  If there were testing effects (citizens were not surveyed; we used police reports), both groups should have reacted the same to an awareness of being studied.
  Thus, the relationship between X and Y is not due to Z [we have gained internal validity (accuracy)].
- Problem: Can we generalize this finding (foot patrol reduces crime) to all police departments in the country? This is the problem of external validity.
- One Solution: Replications (repeat studies) in other settings with other police departments.

FIGURE 3.1 Research Design in a Nutshell.

CAUSALITY

The ultimate purpose of all scientific investigation is to isolate, define, and explain the relationship between key variables in order to predict and understand the underlying nature of reality. The problem of *causality* has been a subject of continuing philosophical discussion, but scientific investigation is based on the a priori assumption that the fundamental nature of reality can be known—that causation lies at the basis of reality.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

Resolution of the Causality Problem

Resolution of Causality Problem

To approach this matter, scientific investigation basically entails *three essential* **Steps for Resolving the Causality Problem**. *The first step* involves the demonstration of a relationship or covariance between variables. That is, one variable is related, increases or decreases in value, in some predictable manner along with increases or decreases in the value of another variable. *The second step* consists of specifying or indicating the time sequence of the relationship. Which variable is the independent or predictor variable *X*, and which is the outcome or dependent variable *Y?* Generally, logic or knowledge of which variable comes first gives one the direction of causation. For instance, it would make more sense to assume that criminality of parents (*X*) would precede in time and possibly predict criminality in offspring (*Y*), rather than vice versa. In most instances, one has little difficulty in identifying the outcome (*Y*) one is interested in predicting. It is usually the subject of the study. Although this process of causality resolution in research has been greatly oversimplified for presentation purposes, most studies of an empirical and predictive nature in criminal justice can be found to undergo essentially the first two steps that have been outlined. *The third step* is the stage where many studies bog down and where research findings are subject to interminable debate. It involves the exclusion of **rival causal factors**, or the elimination of other variables that could conceivably explain away the original relationships the researcher had claimed. Other variables or rival causal factors may be responsible for the variations discovered (Hirschi and Selvin, 1966).

Rival causal factors

Spuriousness

In excluding rival causal factors, researchers are attempting to demonstrate that the relationship between *X* and *Y* is nonspurious. A **spurious relationship** is a false relationship; that is, one that is not caused by the believed variables but can be explained by other variables. The presumed relationship between foot size (*X*) and intelligence (*Y*) may disappear (be demonstrated to be spurious) when controlled for age (*Z*). That is, among thirty-year-olds there is no relationship between foot size and intelligence.

To summarize and clarify this process, once again, the *three essential steps in resolving the causality problem* are:

1. Demonstrate that a relationship exists between the key variables.
2. Specify the time order of the relationship.
3. Eliminate rival causal factors.

Suppose a researcher wanted to prove that a relationship exists between the increase in foot patrols in a precinct and a decline in crime. Assuming that foot patrols have been increased in the target precinct, the researcher looks at some measurement, such as precinct records of reported crime. If no relationship between increased foot patrols and crime is discovered, the entire process stops with step 1. There is little need to proceed if no relationship exists at all. If a decrease or, for some reason, an increase in crime is discovered, however, the researcher goes on to the next step. For our purposes, we will assume that an increase in foot patrols is the predictor variable, *X*, and a decrease in reported crime is the outcome, *Y*. We assume that foot patrols affect crime rates, rather than vice versa. One could see, however, where a researcher might be interested in studying the latter; that is, high crime areas may more likely precipitate the deployment of foot patrols.

Finally, suppose a relationship was discovered and specified: namely, there was an increase in foot patrols and a decline in reported crime within the precinct. Does this then prove that increases in foot patrols will cause a decrease in crime as measured by crime reported to police? The answer to this question is no. The most obvious reason, which will become clear in Chapter 13

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

when we discuss statistics, is that *correlation or relationship by no means implies or demonstrates causation*. Such a finding merely brings us to stage 1 or 2 of our steps in resolving the causality problem. If the prudent investigator has not already guessed, one's critics will very quickly point out that other variables could have accounted for this relationship.

RIVAL CAUSAL FACTORS

Rival causal factors are any variables other than *X* (the treatment) that may be responsible for the relationship. It is traditional in the social sciences, following the lead of Campbell and Stanley (1963), to discuss these other variables, or rival causal factors, as being of two general types: *internal factors* or other variables within the study itself that may tend to invalidate one's findings and conclusions, and *external factors* or elements outside of one's immediate study that may imperil the researcher's attempts to draw generalizations from the study and infer one's findings to be true of larger populations.

Validity

Validity refers to accuracy or correctness in research. Internal factors question the internal validity of research, whereas external factors impugn the external validity of findings. The former asks whether the observational process itself produced the findings; the latter is concerned with whether the results were unique and applicable only to the group or target studied. In checking **internal validity**, one is concerned with whether a variable other than *X* (the treatment) may have produced a change in *Y* (the dependent variable). With **external validity**, one asks what other variables may limit one's ability to generalize the findings in a study to larger populations or settings.

Internal validity

External validity

INTERNAL FACTORS: VARIABLES RELATED TO INTERNAL VALIDITY

Campbell and Stanley's classic monograph, *Experimental and Quasi-Experimental Designs for Research* (1963), points to the following internal factors as possibly threatening the internal validity of a study:

History	Statistical Regression
Maturation	Selection Bias
Testing	Experimental Mortality
Instrumentation	Selection–Maturation Interaction

All of these are rival causal factors that could have been responsible for producing the results rather than the treatment assumed to be responsible. That is, although *X* and *Y* are related, the real reason for this relationship is *Z*, some other variable or rival causal factor.

History

History

History refers to other specific events that may have taken place during the course of the study and may have produced the results. For example, other than increased foot patrols what events may have occurred in the hypothetical precinct and accounted for the decrease in crimes reported? Perhaps the area was the target of a "Crime Watch" program that encouraged citizen vigilance and reporting; or a new employment program was initiated to hire unemployed

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

youths; or urban renewal changed the nature of the population inhabiting the precinct. Social, seasonal, and other events may be responsible for changes in a study target. Garwood (1978) gives an example of a burglary reduction program using "operation identification" in which belongings are engraved to discourage burglars and fences. A northeastern experimental city received the treatment, operation identification, and, especially during January 1978, showed a dramatic decrease in reported burglaries. Can we assume that the *X*, operation identification, was responsible for bringing this about? No. On further investigation, it was discovered that other cities without such a program, notably Buffalo, Detroit, and Boston, demonstrated equally impressive declines in burglary. What may have occurred in January 1978 to account for this? It was the winter of incredibly deep, record-breaking, crippling snows in the Northeast, and thus bad weather, a historical hidden variable, was most likely responsible for the decline in burglary statistics. Similarly, a tough 1880 anti-horse theft law in New York City is not responsible for the virtual disappearance of such thieves a hundred years later; rather, it is the historical change in transportation.

In reviewing shortcomings of sentencing research, Farrington (1978) points out that because of the lack of premeasures and control groups, it is difficult to assume that any decrease in crime after a change in legal penalties or sentencing is due to these changes or other unmeasured social changes that may have taken place at the same time.

Maturation

Maturation

Maturation refers to biological or psychological changes in the respondents during the course of study that are not due to the experimental variable. "Time heals all wounds," according to the old medical dictum, refers to the phenomenon in medical research wherein a given number of patients can be expected to reveal improved conditions with or without treatment. Perhaps the precinct under investigation in our foot patrol example was in the process of change, either deterioration or upgrading, that brought about the change in crime reporting irrespective of foot patrol.

A hypothetical example may serve to illustrate maturation as a rival causal factor. An interesting controversy of the 1960s was fluoridation of water in the United States. Opponents claimed that the addition of such chemicals was potentially harmful. Suppose an avid supporter of such a view were to state, "In 1850 Erie, Pennsylvania, was the first city to fluoridate its water and not a single citizen from that time is alive today." Quite obviously, the demise of this population was due primarily to natural causes, maturation, rather than the assumed cause, fluoridation.

In a more serious vein, claims as to the long-term effectiveness of rehabilitation programs must certainly control for the fact that as a given age cohort matures, its crime commission in general tends to decrease; that is, there are very few eighty-year-old cat burglars. As a more detailed example will illustrate later in this chapter, all other things being equal, older delinquents can be expected to show lower crime commission over time than younger delinquents.

Testing

Testing

Testing (pretest bias) refers to the bias and foreknowledge introduced to respondents as a result of having been pretested. On a second testing, the respondents are no longer naive regarding the subject matter and can make use of sensitivities, information, and attitudes garnered from the first testing. If one wanted to test a bank's reaction to a simulated robbery, the reaction of a bank that had been held up the previous day would probably not yield valid or typical results.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

Instrumentation

Instrumentation

Instrumentation involves changes in the measuring instrument from the beginning or first period of evaluation to the second, later, or final evaluation. The measuring instrument may refer to observers, questionnaires, interviews, analyses of existing records, or any standard method of data gathering. Suppose, for instance, that in our foot patrol experiment, the method of recording citizen complaints was dramatically improved from what it had been at the beginning of the project. An increase in crimes reported to police at the end of the study could very well have resulted from instrumentation, a rival causal factor, rather than the assumed predictor variable—foot patrol. A major limitation of comparing crime rates of today with those of yesteryear relates to the continual improvement in record keeping so that an indeterminate proportion of the increase may simply be evidence of improved instrumentation. We will discuss this subject in detail in Chapter 4 in the section on the Uniform Crime Report. The crime rate of a city may show an increase of 100 percent over the previous year, not necessarily because of increased crime commission, but because of installation of a computer, a change in the manner in which crime is measured or recorded.

Statistical Regression

Statistical regression

Statistical regression is the tendency of groups that have been selected for study on the basis of extreme high or low scores to regress or move toward the mean or average on second testing. As in our example, if a precinct were selected for study on the basis of an extremely high or low volume of citizen complaints, it is expected, irrespective of the treatment variable, that the second reading will be closer to the average for all precincts or certainly less at the extreme.

It should not be surprising that extremely high or low scores would move toward more normal scores upon retest. Extremely tall people as a group are likely to have children shorter than themselves, just as extremely short people are likely to have children taller than themselves. Imagine taking the first examination in a class that you detest; after having had a bad night the night before (illness, an all-nighter), you flunk the test. As a member of the lowest group, you are then chosen for study. But before the second test you have a normal night's sleep, so we might expect your performance to improve. The point is that the improvement may not be due to any increase in intelligence but that your first test performance was atypical.

In critiquing reported positive claims of a program involving diversion alternatives for youths who would otherwise have been incarcerated, Gordon and associates point out that regression effects had been overlooked. The juveniles studied were chosen on the basis of extremely high crime commission that could be expected to decrease upon second observation even without intervention. Not surprisingly, it was claimed that the most extreme delinquents demonstrated the greatest drop in recidivism as a result of a wilderness program. The most likely explanation, however, was an expected regression toward the mean on the basis of the initial choice of extreme cases (Gordon et al., 1978).

Selection Bias

Selection bias occurs when the researcher chooses nonequivalent groups for comparison. Studies that compare the attitudes or behavior of volunteers and nonvolunteers are often subject to selection bias. If in our foot patrol study the precinct chosen for the experiment was characterized by high levels of citizen involvement and reportage of crime, and these data were compared with those for another nonfoot patrol precinct with historically low levels of reported crime, the rival causal factor, selection bias, rather than foot patrol, might explain the differences in findings.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

Similarly, comparison of an experimental group consisting of all model prisoners and a control group of incorrigible prisoners would hardly be fair. Many demonstration projects have been accused of "creaming clients"—taking the cream of the crop—or stacking the deck by assigning the best clients to the treatment and the dregs to the control group.

Experimental Mortality

Experimental mortality

In studying the same group over a period of time, an expected loss of subjects can be anticipated. This loss is referred to as **experimental mortality**. In our foot patrol experiment, a decline in residential population as a result of urban renewal would certainly impinge on the number of crimes reported. In correctional research, long-term recidivism studies have been handicapped by the inability to follow all or most of the original respondents. Perhaps those who cannot be found are more likely to be successes or failures than those on which data are available.

One method of assessing possible bias as a result of the loss of respondents is to compare known characteristics of respondents with those of nonrespondents. Similarity in such demographic characteristics as sex, age, race, and income may lead one to suspect that nonrespondents do not differ much from respondents and therefore introduce little bias.

Selection–Maturation Interaction

Selection–maturation interaction

Obviously, factors within the experiment other than the assumed predictor variable may be responsible for the findings. Interaction or combination impacts of any or all of these variables may bring about the obtained results, for instance, interaction of selection bias and differential maturation of groups. **Selection–maturation interaction** was illustrated by Gordon et al. (1978) in their critique of a diversion program in which the researchers failed to control for age—selection bias—or to spot a potential maturation effect when older delinquents were placed in the wilderness programs that showed the greatest decline in recidivism. The latter can be viewed as a maturation effect in that as a group ages, its overall crime commission declines.

EXTERNAL FACTORS: VARIABLES RELATED TO EXTERNAL VALIDITY

External factors refer to rival causal factors that negatively affect external validity or the representativeness or generalizability of study findings to larger populations beyond the group studied (Campbell and Stanley, 1963). The following are examples of external factors:

- Testing effects
- Selection bias
- Reactivity or awareness of being studied
- Multiple-treatment interference

Although a clever researcher may do much to control for the effects of rival causal factors within a study, this may not enhance the ability to generalize beyond the group studied. The testing effects and selection bias previously discussed as affecting internal validity also affect external validity (reactive or interaction effects of pretesting).

Testing Effects

Testing effects

Testing effects point to the tendency of pretests to destroy the naiveté of respondents with respect to the variable(s) being studied and decrease or more predictably increase the subjects' awareness

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

or sensitivity, thus complicating the ability to generalize their responses to a larger population that has not been pretested. For illustrative purposes, let us alter the foot patrol example by adding a different dimension. Assume that the purpose of introducing walking beats was to enhance community relations and public attitudes toward the police. At the beginning of the study, residents of the precinct were questioned regarding these matters, then foot patrols were introduced, and the residents were questioned again. A more favorable attitude toward police would be assumed to have been produced by walking patrols; however, perhaps it was induced in part, or primarily, by the group having been pretested and thus having had time to reflect and consider their views. Furthermore, attitudes in this precinct could not be generalized to other even similar precincts, without some hazard, unless a similar pretest–posttest had occurred there also.

Selection Bias

Selection bias can have negative impacts on the ability to infer findings beyond the group studied. Nonrepresentative selection of a study group obviously invalidates any attempt to generalize to larger populations. For instance, the purposive selection of a precinct with high citizen vigilance in responding to crime as the setting for an experiment in police deployment would not be a fair test of how this same program would operate in more typical settings.

Reactivity or Awareness of Being Studied

Reactivity

Reactivity or respondent awareness of being studied tends to produce atypical or unnatural behavior on the part of subjects. Most people have had experience with previously announced inspections, visiting or guest teachers and the like, to realize that behavior observed during that day tends to be at times quite different from what normally occurs. Similarly, if the sample foot patrol precinct were announced and continually covered in the media during the course of the experiment, the residents' behavior, as well as the behavior of the police, would be different than usual.

This phenomenon is variously described as the "Hawthorne effect," "placebo effect," or "stooge effect" and will be discussed shortly. Thus, an awareness of being studied, rather than the experimental treatment, may become the major factor bringing about a particular outcome.

Multiple-Treatment Interferences

Multiple-treatment interference

Multiple-treatment interference occurs when more than one treatment or predictor variable is used on the same subjects. The outcome may be brought about by a specific sequence or combination of independent variables that can be uncovered only by more complicated research designs, as will be examined later in this chapter. If the foot patrol officers also wore blazers and did not carry guns and gave out free tickets to sports events, a more positive attitude may have been produced by any one or combination of variables, in addition to or irrespective of foot patrols.

RELATED RIVAL CAUSAL FACTORS

Hawthorne Effect

Hawthorne effect

Although not distinct from those already discussed, a number of other Hawthorne-related terms for sources of invalidity can be identified. The **Hawthorne effect** serves as an example of reactivity resulting in atypical behavior or attitudes on the part of research subjects as a result of their awareness of being studied. This factor gets its name from a pioneering industrial study

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

of a group of workers in the Hawthorne plant of the Western Electric Company in Chicago (Roethlisberger and Dixon, 1939). To the bemusement of the researchers, alterations in treatment designed to either increase or decrease worker efficiency consistently increased efficiency. Rather than reacting to the treatment variable, *X*, workers were reacting to a rival causal factor—the fact that they had been singled out for an important study. The workers reacted as they suspected the researchers wanted them to act, rather than as they would under normal circumstances. These acquiescent, "guinea pig," or "stooge" effects are likely in situations where the group being studied is aware that they are being studied.

Halo Effect

Halo effect

The **halo effect** was coined by Thorndike (1920, p. 25), who noticed that when supervisors rated subordinates, the ratings were all "higher than reality." It refers to observer bias in which observers, perhaps unconsciously, follow an initial tendency to rate certain objects or persons in a particular manner; this initial orientation carries over into all subsequent ratings. The less specified and discretionary the variable to be rated, the greater the danger of the halo effect (see Cooper, 1981).

Self-fulfilling prophecy

Related in part, but more subtle than the halo effect, is the carryover into research of a phenomenon first noted by sociologist W. I. Thomas. His basic maxim of "the definition of the situation," or what others refer to as **self-fulfilling prophecy**, has a major bearing upon the bias of the researcher. "If groups or individuals define a situation as real, it is real in its consequences" (Thomas and Swaine, 1928, pp. 571–572). A researcher's own hidden biases and expectations may influence his or her perception of events so as to bring about that which had been assumed. Selective perception may lead to one ignoring anything that does not fit one's preset cognitive map and thus presents us with an experimenter effect (Rosenthal, 1966).

Post Hoc Error

Post hoc error

Post hoc error comes from the Latin phrase "post hoc; ergo, propter hoc," literally, "after this; therefore because of it." It is a fallacy to argue that one variable is the cause of an outcome because it precedes that outcome in time. What is considered an effect is often only a subsequent event. An example would be to argue that because every morning when the rooster crows the sun rises—the crowing causes the sunrise. Another example can be the common claims made by police chiefs that crime (reported crime) declined in their city in the 1980s because of new and effective policies. More likely than not, what one was observing was the predicted "crime dip" of the 1980s, resulting in part from demographic shifts in what has been described as the post–World War II "baby boom."

Gelles (1977) gives an example of such fallacious reasoning in research on child abusers. Sometimes psychological conditions that are identified as being present after the abuse incident are viewed as the cause of the incident; for example, abusers may be found to be paranoid and depressed, conditions that may be results of the incident, rather than its cause.

Placebo Effect

Placebo effect

Complete enumeration of related tags or descriptions of rival causal factors would be endless and beyond our needs at this point. One final factor that often appears in literature is the **placebo effect**. This involves, similar to the Hawthorne effect, the tendency of subjects to react to a known stimulus in the predicted manner (Loranger, Prout, and White, 1961). Commonly used in medical research, the "sugar pill" or a placebo (fake treatment) with no known effects is administered to the study group to hide the real treatment group and also to control for the placebo effect. The actual effects of the true experimental pill can then be compared with effects induced in the

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

control group by the sugar pill. Other terms used to refer to this same reactivity phenomenon are *evaluation apprehension* or *demand characteristics*. In the former, Rosenberg (1969) notes that subjects are apprehensive about participating in experiments, and this very anxiety may produce atypical behavior. "Was I a good subject?" is often the question asked by respondents who are willing to, in an acquiescent manner, elicit on demand characteristics they believe are sought by the researcher (Orne, 1969).

Researchers have found that subjects on placebos who experience pain relief actually have their own brains releasing pain-relieving chemicals called endorphins (Kaptchuk, Eisenberg, and Komaroff, 2002). The brain's amazing triumph over reality can be illustrated by studies in which hair growth was maintained or improved in 42 percent of balding men taking a placebo; smelling a placebo helped asthmatic children in Venezuela increase their lung function by 33 percent; and in a Japanese study, people exposed to fake poison ivy developed real rashes (Blakeslee, 1998).

Double-blind experiment

To control for the "placebo," "Hawthorne," and "experimenter" effects, medical researchers have developed the gold standard for experimental research, the **double-blind experiment**, a design in which neither subjects nor people administering the experiment are aware of which group is the experimental group (the one receiving "real" treatment) and which group is the control group (the one receiving "phony" treatment) (Glaser, 1976, p. 773).

EXPERIMENTAL DESIGNS

Research designs

Research designs are a major way of controlling for invalidity in research or, as in step 3, resolving the causality problem, a means of excluding rival causal factors (Cresswell, 1994). Previously in this chapter, we indicated that experiments may be viewed as the benchmark for comparison of all other research methods. The language of research design is heavily couched in that of experimental designs, and a mastery of this specialized jargon is helpful in developing a standard taxonomy by which to classify various forms of research.

Three general types of experimental designs are discussed in this section:

1. *Experimental designs* (sometimes called "true experimental designs") are characterized by random assignment to treatment and control groups and include the classical, posttest-only control group, and Solomon four-group designs.
2. *Quasi-experimental designs* do not use random assignment of groups and instead employ matching or other means of obtaining equivalence of groups. Quasi-experimental designs include time-series and counterbalanced designs.
3. *Preexperimental designs* lack any equivalence of groups and include one- and two-group ex post facto and one-group before–after designs.

What we will describe as the classic experimental design is one of the most effective methods of controlling for internal rival causal variables before the fact. In that sense, the experiment becomes the point of departure for a comparison of all other research designs.

THE CLASSIC EXPERIMENTAL DESIGN

Classic experimental design

The **classic experimental design**, which serves as a prototype for our discussion of all other research designs, *contains three key elements*:

Equivalence

Pretests and posttests

Experimental and control groups

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

Equivalence

Equivalence refers to the attempt on the part of the researcher to select and assign subjects to comparison groups in such a manner that they can be assumed to be alike in all major respects. The two methods by which equivalence of groups to be compared is gained are randomization and matching. **Randomization** is the random assignment of subjects from a similar population to one or the other group(s) to be compared in such a way that each individual has an equal probability of being chosen and an equal probability of being assigned to any of the groups to be compared. We discuss the process of randomization more thoroughly in Chapter 4, but at this point, it will suffice to indicate that one of the principal means of accomplishing randomization is by use of simple random samples or some means of selection in which each case in the population has an equal probability of appearing.

Randomization

Matching

Matching deals with assuring equivalence by selecting subjects for the second or other comparison groups on the basis of matching certain key characteristics such as age, sex, and race, so that the groups are similar or equivalent with respect to these characteristics. Matching and randomization can be combined. In the Cambridge–Somerville study (McCord and McCord, 1959), 325 pairs of boys were matched on delinquency potential, and one member of each pair was randomly assigned to treatment (Farrington, 1983, p. 261).

Pretest–posttest

Experimental group

Control group

Assuming that the groups are similar, both are exposed to a **pretest** or observation prior to exposure to treatment and a **posttest** or measurement after exposure to treatment. Finally, the group exposed to treatment is called the **experimental group**; the group that is not exposed to the stimulus or predictor variable is the **control group**. The original meaning of the term “control” is “check”—the term comes from “counter-roll,” a duplicate register or account made to verify an account (Oakley, 1998).

For heuristic purposes, we adapt the notation developed by Campbell and Stanley (1963) for schematically depicting the various research designs: *X* equals treatment, *O* symbolizes observations (some researchers use *T* instead of *O*). Subscripts for *O*, such as O_1 and O_2, represent the first and second observations, respectively, and *E* stands for the equivalence of comparison groups. Please note that purists would be more conservative and insist that randomization (*R*) and not matching is necessary for a classic experimental design.

Classic Experimental Design

$$\begin{array}{cccc} E & O_1 & X & O_2 \\ E & O_1 & & O_2 \end{array}$$

E = equivalence
O = observation
X = treatment
1, 2 = time

Following our previous presentation of this design with our newly introduced notation, we find that the classic experimental design involves an equivalent assignment to experimental and control groups, which are observed both before and after the experimental group receives treatment.

Using some of the rival causal factors affecting internal validity, we can now examine the point that experimental designs are effective in controlling for many of these sources of error before the fact. Suppose a classic experimental design had been employed for the foot patrol experiment discussed earlier. Two precincts alike in all possible respects would have been chosen for study (equivalence). The experimental precinct would have been pretested prior to the treatment (foot patrol), whereas the control precinct would have received no treatment (retained usual

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

patrol practices). At the end of the study, both groups would once again have been observed, and any differences between them would be assumed to have been produced by the deployment of foot patrols. History and maturation would not be likely rival explanations for these differences, because both groups were similar and thus exposed to the same historical and maturation conditions. Both groups were exposed to a pretest; thus, the differences could not result from the pretest, or at least the extent to which the pretest influences results can be assessed because it will show up in both groups. If it is assumed that both groups received the same instrument, changes in the instrument would not account for differences as long as these changes were the same for both groups. Statistical regression should be the same for both groups. Equivalence assures no selection bias and experimental mortality should be the same.

This hypothetical foot patrol study is an example of a *field experiment*, a type of experiment that is conducted in a natural (field) setting; a *laboratory experiment* is a type of experiment that takes place in contrived or researcher-created conditions. Our next examples illustrate field experiments. Because most of these studies employ matching rather than randomization as the means of assuring equivalence, purists might argue that they are not true experimental designs but rather quasi-experimental designs.

SOME CRIMINAL JUSTICE EXAMPLES OF THE CLASSIC EXPERIMENTAL DESIGN

Candid Camera

While the use of videotape for security purposes is routine today, at one time the usefulness of such electronic devices (in this case, photographs) was not commonly understood. In 1975, the Seattle Police Department installed hidden cameras in stereo speaker boxes in seventy-five commercial establishments identified as high-risk potential robbery victims. These businesses constituted the experimental group; a group of similar businesses received no treatment. The pretest for both groups consisted of gathering statistics on the percentage of robberies cleared by arrest and conviction rates prior to the study. If held up, a clerk triggered the camera by pulling a "trip" bill from the cash drawer. A special project director would make prints of the photograph of the robber available. A posttest comparing the two groups found that 55 percent of the robberies of experimental companies were cleared by arrest versus 25 percent of the control firms. Similarly, 48 percent of robbers at hidden camera sites were convicted, compared with 19 percent of the control group robbers ("Hidden Cameras Project," 1978).

Scared Straight

In the late 1970s, claims of reduced recidivism among juveniles in trouble with the law in response to visits with prisoners were illustrated by the Rahway prison project portrayed in the film *Scared Straight*. The assumption was that much of the "glamor" attached to criminal life by juveniles on the road to more trouble could be nipped in the bud by blunt, heart-to-heart dialogue with specially selected prisoners. Many jurisdictions began to set up what appeared to be a new gimmick in corrections. Later evaluations, however, suggested that the benefits claimed were premature.

Yarborough conducted an evaluation of the JOLT (Juvenile Offenders Learn Truth) program at the State Prison of Southern Michigan at Jackson. Unlike some earlier programs, verbal attacks and obscenities were deemphasized. In 1978, subjects were randomly assigned to experimental and

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall. Copyright © 2010 by Pearson Education.

control groups and then measured at three- and six-month follow-up periods. All subjects were male and had been arrested or petitioned for an offense that, had they been adults, would have been criminal. No significant differences were found between those who had attended the JOLT session and those who had not. There were no differences in the proportion having petitions filed or in the types of offenses committed (*Scared Straight*, 1979).

In 1987, WOR TV (Secaucus, New Jersey, June 21, 9:00 P.M. EST) aired a documentary entitled "*Scared Straight*: Ten Years Later." This program presented a very positive picture of the experiment, relying primarily on interviews with people who had attended. Finckenauer (1982), in *Scared Straight! and the Panacea Phenomenon*, felt that the original documentary (1977) had misled the American public into thinking this was a miracle cure for juvenile crime. Despite many problems and lack of cooperation in attempting to evaluate the "Juvenile Awareness Program" or "Lifer's Program" (*Scared Straight* Program) at Rahway, Finckenauer found that his randomly assigned *Scared Straight* experimentals actually had higher seriousness delinquency scores afterward than the controls who did not attend the program. Similarly, many of the juveniles put through the program were not the hardened "junior criminals" the public had been led to believe.

A review of nine *Scared Straight* experiments over its thirty-three-year history indicated that, despite good intentions, the program not only did not deter future criminal behavior but actually led to more crime by program participants. The researchers concluded: "Given the possibility of harmful effects of interventions, government has an ethical responsibility to rigorously evaluate, on a continuing basis, the policies, practices, and programs it implements" (Petrosino, Turpin-Petrosino, and Finckenauer, 2000).

Community Policing

Community policing has become a subject of much interest in law enforcement since the 1980s. The term broadly refers to a variety of strategies that attempt to get the police away from rapid response to service and closer to the community on a day-to-day basis. Order-maintenance, community crime prevention, problem solving, neighborhood safety, foot patrol, and a host of police–community relations strategies are all included under community policing (Mastrofski, 1992).

A variety of field experiments regarding the impact of neighborhood safety programs such as "neighborhood or block watch," "police storefronts," and "foot patrol" experiments have been undertaken. The Police Foundation's analysis of a neighborhood watch experiment in Houston found no noticeable reduction in crime compared with a similar area that had not received the program. Another Houston program established a police department storefront (a combination precinct station, social center, and community outreach center). Houston also experimented with personal-contact patrol in which officers attempted to stop and talk to as many citizens as possible. Despite problems in maintaining experimental conditions, some positive findings were obtained. The neighborhood watch and storefront programs had no noticeable impact upon crime reduction but an enormous impact upon reduction of citizens' fear of crime. In addition to reducing fear, the personal-contact patrol reduced household victimizations by one-half and resulted in improving the attitudes of residents on community issues (Sherman, 1985; Wycoff et al., 1985a, 1985b). Exhibit 3.1 describes a field experiment in Kansas City that met with success.

A continuing subject of debate has been public dissatisfaction with anonymous, routine policing in automobiles and requests for foot patrols, even though most police managers until recently viewed such assignments as inefficient deployment of limited police personnel.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

EXHIBIT 3.1

The Kansas City Gun Experiment

The United States has both the highest violent crime and homicide rate of any developed country, as well as the largest armed civilian population in the world. National attempts to significantly control firearms are effectively blocked by the powerful National Rifle Association. Given these contingencies, what can police do to try to control growing youth homicide rates? One possibility tested in the Kansas City gun experiment by Sherman, Shaw, and Rogan (1995) was that greater enforcement of existing laws against carrying concealed weapons could reduce gun crime. With a Bureau of Justice Assistance "Weed and Seed" program grant, the Kansas City Police Department selected a target patrol beat and a control beat. The target beat had a 1991 homicide rate of 177 per 100,000 persons, about twenty times the national average. The control beat had a similar violent crime rate. The research design involved a matched groups before–after design. The "hot spot" target area received increased proactive patrols.

> The actual technique the officers used to find guns varied, from frisks and searches incident to arrest on other charges to safety frisks associated with car stops for traffic violations. Every arrest for carrying concealed weapons had to be approved for adequate articulable suspicion with a supervisory detective's signature (ibid., p. 6).

Figure 1 illustrates the differences between the target beat and the comparison beat during the one-year experiment.

Gun crimes in the target beat decreased from 37 per 1,000 persons to 18.9 and guns seized increased from 9.9 to 16.8. The comparison beat showed little change in gun crime (22.6 to 23.6 per 1,000) and an actual decrease in guns seized (10.4 to 8.8 per 1,000). There was no displacement of gun crimes to surrounding areas. Drive-by shootings dropped from 7 to 1 in the target area but doubled

HAMILTON, STEVEN 4228BU

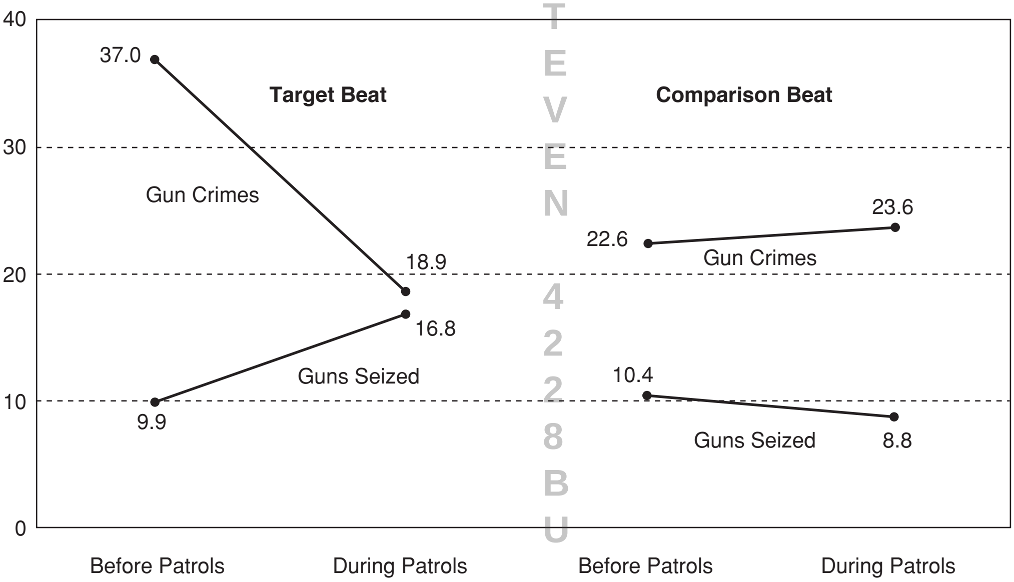

FIGURE 1 Kansas City Gun Experiment—Guns Seized per 1,000 Persons. (*Source*: Sherman, Shaw, and Rogan, 1995, p. 1.)

(*continued*)

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

EXHIBIT 3.1 (Continued)

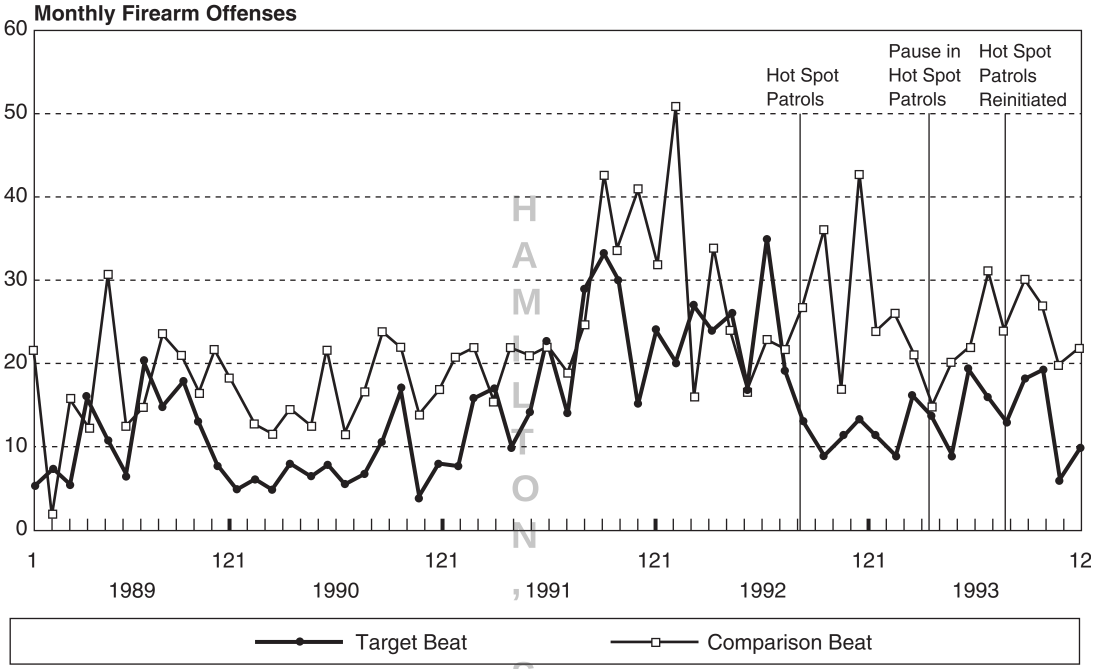

FIGURE 2 Total Offenses with Firearms by Month in Target and Comparison Beats. (*Source*: Sherman, Shaw, and Rogan, 1995, p. 7.)

from 6 to 12 in the comparison area, again with no displacement effect. Figure 2 compares offenses by firearms.

Other positive findings were a decline in homicides in the target area but not in the comparison area. Citizens in the target area were less fearful, but there was no change in fear in the comparison area. Two-thirds of those arrested for gun carrying in the target area were from outside the area.

Finally, only gun crimes were reduced by the directed patrols, and they had no effect on calls for service or reduction of other crimes. Further replications are underway as of this writing. Lawrence Sherman, the principal author of the report, is on leave from the University of Maryland and is serving as Criminologist to the Indianapolis Police Department in order to further test the program.

Source: Sherman, Lawrence W., James W. Shaw, and Dennis P. Rogan. "The Kansas City Gun Experiment." National Institute of Justice Research in Brief, January 1995. Document available from National Criminal Justice Reference Service, Box 6000, Rockville, MD 20849-6000; call 1-800-851-3420 or Internet lookncjrs@aspensys.com.

HAMILTON, STEVEN 4228BU

During the late 1970s, experiments with foot patrols were conducted in Newark, New Jersey, and Flint, Michigan (Police Foundation, 1981; Wilson and Kelling, 1982; Trojanowicz and Banas, 1985). The same findings were obtained in both studies (Kelling, 1985, p. 2): decreased fear of crime, greater citizen satisfaction, and greater appreciation of neighborhood values by the police. There also appeared to be greater job satisfaction, less fear, and higher morale for officers who patrol on foot than for officers who patrol in automobiles. The

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

Flint study showed a decrease of 40 percent in service calls via telephone and a modest reduction in crime, whereas the Newark program showed no crime reduction. More replications (repeats of the experiments) are under way. Foot patrols are obviously no panacea but have been found popular, particularly when selectively implemented in densely populated urban areas.

In the 1990s, the National Institute of Justice had contracted a large number of studies of community policing ("Community Policing," 1992). For example, an evaluation of the Madison, Wisconsin, Police Department's "quality" policing program examined a community-oriented policing program with a new organizational design based on the work of management expert Edwards Deming. In comparing the experimental with the comparison police district, it was found that in the experimental district the managers viewed themselves more as problem solvers and employee attitudes regarding their work and organization improved. In addition, citizen interaction improved and their perception of crime as a problem was reduced. Citizens also expressed a more positive attitude toward the police (Wycoff and Skogan, 1994).

OTHER EXPERIMENTAL DESIGNS

Other experimental designs include the *posttest-only control group design* and the *Solomon four-group design*.

Posttest-Only Control Group Design

To assess the impact of consolidated police departments on public satisfaction with policing in their area, Ostrom, Parks, and Whitaker conducted a *posttest-only control group design* (1973).

$$\begin{array}{lll} E & X & O \\ E & & O \end{array}$$

Three Indianapolis neighborhoods with a consolidated police department were matched with three communities with independent police departments. The researchers discovered higher public satisfaction in the communities with independent departments. Without a pretest it is unclear, however, whether this difference could not in fact have been caused by other rival factors; for example, perhaps even before police department consolidation, the Indianapolis neighborhoods had lower degrees of satisfaction with their policing.

Solomon Four-Group Design

The *Solomon four-group design* (Solomon, 1949) is viewed by some as the purest of research designs. Basically, it combines the classic experimental design with the posttest-only design. The Solomon design has four groups, the first two resembling a classic design and the second two resembling a posttest-only control group design.

$$\begin{array}{llll} E & O_1 & X & O_2 \\ E & O_1 & & O_2 \\ E & & X & O_2 \\ E & & & O_2 \end{array}$$

HAMILTON, STEVEN 4229BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

To illustrate the Solomon design, let us hypothetically suppose, using the same Indianapolis example, that four areas could be chosen for study from evenly matched communities. The first two areas would both be measured with respect to public attitude toward the police. Only one of these would receive the treatment (consolidated policing), and both would be remeasured with respect to attitude. Two other areas would not receive the premeasure; that is, they would be measured only after one had received treatment (consolidated policing) and one had not. Such a design would assess the effect of testing effects as well as provide a premeasure lacking in the posttest-only control group design. The advantage of the posttest-only control group design is that it eliminates testing effects and possibly reactivity entirely, although it lacks a measure of where the groups stood prior to the treatment. The Solomon four-group design obviously has the advantage of having the premeasure and, by adding the second two groups, has the same advantage as the posttest-only control group design. It is, however, expensive and difficult to implement and, therefore, not practical in many research situations.

The classic Solomon and posttest-only control group designs are examples of experimental designs. Randomization, in which equivalence is obtained by random assignment of subjects to experimental and control groups, is the key distinguishing characteristic of experimental designs. Preexperimental designs lack equivalence of groups, and quasi-experimental designs rely on matching of subjects to achieve equivalence. In many field experiments such as the Indianapolis example, randomization may be inappropriate or impossible; as long as equivalence is assured, it can be argued that these are true experiments.

PREEXPERIMENTAL DESIGNS

Research designs that lack one or two of the three major elements of experimental designs—equivalence or experimental and control groups—are designated as preexperimental designs.

One-Group Ex Post Facto Design

X *O*

One-Group Before–After Design

O *X* *O*

Two-Group Ex Post Facto Design

X *O*

O

One-Group Ex Post Facto Design

All of these preexperimental designs fail to provide equivalence or any assurance that the group(s) being studied is representative in any way of some larger population(s). The *one-group ex post facto design*, or one-shot case study, is quite typical of many early criminal justice demonstration projects. Our original example of the precinct foot patrol experiment, if it contained no preobservation, would serve as an illustration. Unlike true experimental designs, one-group ex post facto (after the fact) studies are subject to many internal invalidity factors or errors. One simply has chosen for study a group that has already been exposed to a particular treatment. Obviously, things other than the treatment could explain the outcome. If one finds a precinct that had experimental foot patrols or an agency with low recidivism rates, without premeasures, equivalence, and control groups, one is on shaky ground in concluding that lower crime rates or lower recidivism are due to these factors. Many studies, particularly field studies in criminology and criminal justice, are of

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

the one-shot case study variety. Cressey's (1957) study of incarcerated embezzlers, for instance, may have major problems with respect to selection bias and reactivity but, on the other hand, may be the only way of obtaining exploratory information on a little-known topic. What one-group ex post facto studies lose in terms of internal control of error may be gained in terms of studying groups in natural field conditions.

Early research on the XYY or "supermale syndrome" assumed, on the basis of studies of incarcerated violent offenders, that an extra male chromosome may have been responsible for violent crime (Witkin et al., 1978). Only later examination of the general population suggested that it may be as prevalent among the "noncriminal" male population. As mentioned previously, in lamenting the shortcomings of studies of sentencing behavior, Farrington (1978) points out that few studies use both before and after measures or compare a sentenced group with an unsentenced group, thus making it difficult to know whether changes in behavior are due to sentences, penalties, or other concurrent social changes.

In another example of a one-group ex post facto study, Heussenstamm (1971) reports on a field experiment in which subjects, none of whom had received traffic violations the previous twelve months, attached Black Panther bumper stickers to their automobiles. They attracted so many traffic citations that the experiment had to be canceled. That Heussenstamm had knowledge of the fact that the subjects had not received citations before the treatment may qualify the study as being an example of the next type, the one-group before–after design.

One-Group Before–After Design

The *one-group before–after design*, or one-group pretest–posttest design, is an example of a longitudinal design. A group, which is not necessarily chosen on the basis of representativeness, is observed, exposed to treatment, and again observed. The primary advantage of this design over the one-group ex post facto design is, of course, the presence of a premeasure. This adds, however, the problem of testing effects and has the same problem as the one-group ex post facto design in that one's findings cannot be compared with those for a similar control group not exposed to treatment.

An example of a one-group before–after design is Pierce and Bowers' (1979) analysis of the impact of the Massachusetts Bartley–Fox gun law, which carried a one-year minimum prison sentence for the unlicensed carrying of firearms. With the use of recorded crime statistics and observations both before and after passage of the law, the earliest part of a longitudinal design suggested a decrease in gun-related assaults, robberies, and homicides; however, this was offset by increases in nongun assaults and robberies using other weapons. Without a control group, the same problem exists as in our previous burglary reduction program (Garwood, 1978)—other variables may be responsible for these findings.

Two-Group Ex Post Facto Design

The *two-group ex post facto design* eliminates possible pretest reactivity by studying both an experimental and a control group after the experimental group has been exposed to some treatment. The primary problem with this design is that there is no way of being sure that the two groups were initially equivalent. Skillful selection of groups for comparison may be the only option a researcher has in some instances. Brown et al. (1970) surveyed two groups of parolees: those who had succeeded and those who had failed at parole. After the fact, they were asked to identify factors in the institution and community that assisted or impeded their

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall. Copyright © 2010 by Pearson Education.

adjustment. Such two-group ex post facto designs were heavily utilized in early biological and psychological theories. These theories in criminology that claimed genetic or personality differences between criminals and noncriminals suffered from what Reid (2005, p. 89) describes as the **dualistic fallacy**, the assumption that prisoners (who are supposed to represent criminals) and groups from the general population (all of whom are assumed to be noncriminals) represent mutually exclusive groups (or are nonoverlapping).

Dualistic fallacy

CROSS-SECTIONAL AND LONGITUDINAL DESIGNS

Before going further, it is crucial to briefly introduce a general distinction used to describe research designs. *Cross-sectional designs* involve studies of one group at one time and usually refer to a representative sample of this group. *Longitudinal studies* are studies of the same group over a period of time and generally are studies of change (Menard, 1991).

Time-series designs involve variations of multiple observations of the same group at various times. Variations may include dividing the original group into equivalent groups and observing these portions longitudinally. In Chapters 5 and 6, we discuss the usefulness of panel designs in such surveys as the National Crime Victimization Survey in providing an in-depth view over time of the same study population. In a now classic criminological study, Wolfgang et al. (1972) used existing records to longitudinally trace the criminal or noncriminal careers of 9,945 boys born in Philadelphia in 1945.

In a replication that included females, Tracy, Wolfgang, and Figlio (1985) tracked the criminal history of males and females born in Philadelphia in 1958 who continued to live there from the age of ten until adulthood. Both studies were instrumental in identifying the concept of serious career criminals, finding that approximately 6 percent of the 1945 group had been responsible for 53 percent of arrests for violent crime and 71 percent for robbery, whereas 7 percent of the 1958 group had committed 75 percent of all serious crime by this group. Farrington (1979) conducted a similar study in London begun in 1961 of boys eight to nine years old in state primary schools.

One of the earliest series of cohort analysis was done by the Gluecks, who studied 500 reformatory inmates over ten years and 1,000 juvenile delinquents for more than fifteen years (Glueck and Glueck, 1937, 1940). Such longitudinal designs are useful in giving us the long- and short-term variations over time. Another example of an ambitious longitudinal study was the Cambridge–Somerville Youth Study begun in 1937 and continuing with some interruption through 1945. Extensive data were gathered on each of the 650 boys who began the project including delinquency, neighborhood, family conditions, school behavior, intelligence, and personality. In 1955, the McCord and McCord (1959) reexamined the data and compared official and unofficial delinquents. Exhibit 3.2 examines a longitudinal study of child abuse and neglect victims. The Denver Youth Survey (Browning and Huizinga, 1999) is another example of a longitudinal study. In order to discover correlates of crime, it has been following 1,527 boys and girls from high-risk neighborhoods in Denver who were seven, nine, eleven, thirteen, and fifteen years old in 1987.

Time-series designs and panel designs are other terms used to refer to types of longitudinal studies, as are cohort and trend studies which are also variations of longitudinal designs. **Time-series designs** involve measuring a single variable at successive points in time. In an *interrupted time-series design*, measurements are taken at time points prior to treatment and for an equivalent period after intervention. The rate of crime committed one

Time-series designs

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

EXHIBIT 3.2

The Cycle of Violence and Victims of Child Abuse

In previous research, Cathy Spatz Widom (1992) found that childhood abuse or neglect increased the odds of future delinquency and adult criminality by an overall 40 percent. The study consisted of a longitudinal study of 1,575 cases from childhood through young adulthood. A group of 908 cases of child abuse or neglect processed by the courts between 1967 and 1971 were tracked using official records for fifteen to twenty years. A comparison group of 667 was matched by age, sex, race, and family social class. The study concluded that:

> While most members of both groups had no juvenile or adult criminal record, being abused or neglected as a child increased the likelihood of arrest as a juvenile by 53 percent, as an adult by 38 percent, and for violent crime by 38 percent (ibid., p. 1).

Using these same cases, Widom (1995) also examined the relationship between childhood sexual abuse and later criminal behavior, particularly sexual offenses. The key finding was that

> People who were sexually victimized during childhood are at higher risk of arrest for committing crimes as adults, including sex crimes, than are people who did not suffer sexual or physical abuse or neglect during childhood. However, the risk of arrest for childhood sexual abuse victims as adults is no higher than for victims of other types of childhood abuse and neglect (ibid., p. 2).

Compared with victims of physical abuse, child abuse victims are more likely to be arrested for prostitution. Victims of physical abuse were more likely to commit rape and sodomy than were sexual abuse victims or the nonvictimized. The long-assumed relationship between childhood sexual abuse, running away, and prostitution was not borne out by the research.

All of these findings relied on official statistics for measuring the dependent variable of crime commission. Continuing research in this series is examining other sources. An attempt is being made to reinterview all 1,575 subjects in order to discover other consequences of child abuse including social, emotional, cognitive, psychiatric, and health outcomes. Also to be examined are factors which protect child abuse victims from later negative consequences.

Sources: Widom, Cathy Spatz. "The Cycle of Violence." National Institute of Justice Research in Brief, October 1992; Widom, Cathy Spatz. "Victims of Childhood Sexual Abuse—Later Criminal Consequences." National Institute of Justice Research in Brief, March 1995. Documents can be obtained from the National Criminal Justice Reference Service, Box 6000, Rockville, MD 20849-6000; call 1-800-851-3420 or Internet lookncjrs@aspensys.com.

year prior to treatment could be compared with the rate for the first year after treatment (Schneider and Wilson, 1978). *Trend studies* analyze different samples of the same general population longitudinally, whereas *cohort studies* analyze subgroups over time, although each time may consist of a sample of the cohort. *Panel studies* examine the same select group or sample over time, as we will see in the National Crime Surveys. The most ambitious longitudinal study ever conducted in the social sciences is the previously discussed "Project of Human Development" of Chicago neighborhoods.

Although there has been some debate in the field of criminology regarding the overapplication of scarce federal research funds to expensive longitudinal studies, it is often the only way of sorting out many trends and causal relationships (Esbensen and Menard, 1990, p. 5).

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

QUASI-EXPERIMENTAL DESIGNS

There are many variations of the experimental design. In fact, as mentioned previously, almost all research in criminal justice can be described using the notation with which we have been working. Because quasi-experimental designs rely on matching—the use of "comparison groups" or means other than randomization to obtain equivalence—the value of using comparison groups depends upon how similar the groups are on key variables to the treatment group. Some quasi- or semiexperimental designs include single time-series, multiple time-series, and counterbalanced designs.

Time-Series Designs

Time-series designs refer to the analysis of a single variable (e.g., crime rate) at many successive time periods with some measures taken prior to treatment and other observations taken after the intervention. It is sometimes called an interrupted time series because the series of observations is interrupted by a treatment (*X*).

Interrupted Time-Series Designs

O O O O X O O O O

It is desirable to have at least ten preobservations and a bare minimum of two, but probably more, postobservations (Schneider et al., 1978, pp. 2–13). Such designs are widely used in criminal justice research in examining the impact of a new law or treatment upon trends in crime.

Interrupted time-series designs can then be defined as an analysis of a single variable measured at many successive time points, with some measures taken prior to a treatment (interruption) and others taken after the treatment. Preproject observations are used as a basis for estimating the trend, and differences between this projected trend and the trend observed after treatment can be assessed to determine whether the treatment had an impact.

Figure 3.2 depicts time-series data for a problem-oriented policing program (the treatment) designed to reduce larcenies from automobiles. The overall reduction in trend lines can be noted from before to after the intervention.

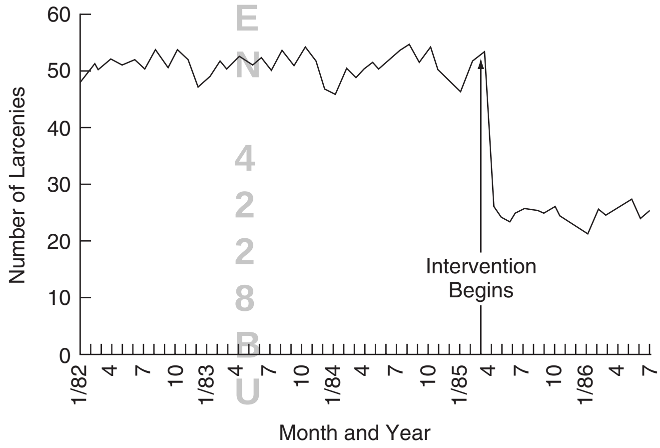

FIGURE 3.2 Time-Series Data for Larcenies from Automobiles in Newport News, Virginia. The Intervention (Treatment) Was a Problem-Oriented Policing Approach That Consisted of Special Tracking and Investigation of Crime Incidents. (*Source*: Spelman, William, and John E. Beck. "Problem-Oriented Policing." *Research in Brief.* National Institute of Justice (January 1987): 7.)

HAMILTON STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

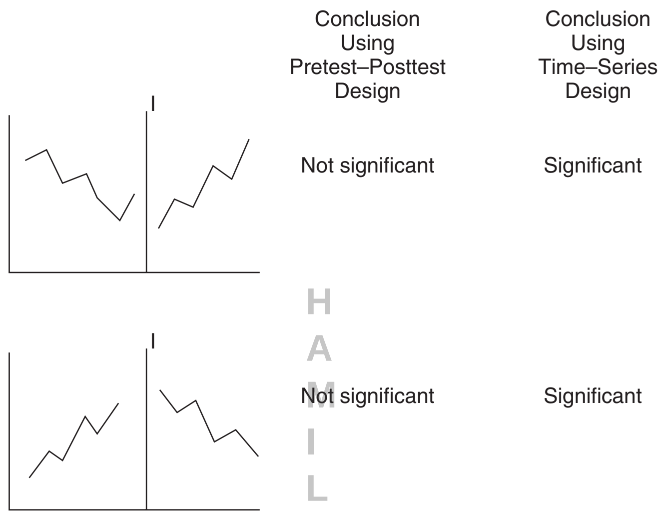

FIGURE 3.3 A Comparison of Pretest–Posttest Designs and Time-Series Designs. (*Source*: Schneider, Anne L., et al.*Handbook of Resources for Criminal Justice Evaluators*. Washington, D.C.: National Institute of Law Enforcement and Criminal Justice (1978): 2–47.)

Figure 3.3 demonstrates the advantage of time-series designs over simple pretest–posttest designs.

In both instances, simple analysis of the last point before and the first point after the intervention would have led to the conclusion of no significant change where, in fact, examination of trend lines showed significant change.

Monahan and Walker (1990, p. 66) give an illustration of the superiority of a time-series design over a before–after design in their analysis of the impact of the Community Mental Health Centers Act of 1963. This program's goal was the reduction of state mental hospitalization.

> In 1963, the year the act was passed, the resident population of state mental hospitals in the United States was approximately 500,000. In 1990, it was less than 150,000. These before–after figures have been used to persuade Congress of the effectiveness of the act. When a time-series with more than one measurement before the passage is used, however, the results seem quite different. A time-series shows the population of state mental hospitals to have increased each year from early in the century until 1955, and decreased each year thereafter, with no noticeable acceleration in the rate of decrease in 1963, the year the act was passed. In this light, the most plausible hypothesis is that the factor causing the population decrease began in the mid-1950's, and not in the mid-1960's. Many now view the introduction of psychotropic medication as the principal method of treating mental patients, which indeed began in 1955, as the most plausible hypothesis to account for the "deinstitutionalization" of mental hospitals.

In yet other variations of interrupted time-series designs, the impact of new prisons (experimental counties) was compared with matched/control counties in order to examine the impact of the prisons. Twenty variables were followed for two years before and two years

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

after the prison openings (Smykla et al., 1984). The relationship between "War and Capital Punishment" was examined by studying the number of executions before, during, and after three war periods (Schneider and Smykla, 1990).

Multiple Interrupted Time-Series Designs

A distinction is also made between *single interrupted time-series designs*, which examine one group or site's preprogram and postprogram outcomes over time, and *multiple interrupted time-series designs*, which contrast one group's performance with that of relevant comparison groups.

Multiple Time-Series Designs

O	O	O	O	X	O	O	O	O
O	O	O	O		O	O	O	O

Although a single interrupted time-series design might examine the impact of the 55-mph speed limit upon vehicular deaths, a multiple interrupted time-series design would compare a state with the new speed limit with one without it during the same period. For example, Connecticut's 1955 crackdown on speeding reduced traffic fatalities, whereas neighboring states without this program experienced no decrease (Campbell and Ross, 1980). In the 1960s, Boston and New York City had very restrictive handgun licensing laws but complained that their laws were defeated by guns brought in from nonrestrictive states. Zimring (1975) studied the impact of the Federal Gun Control Act of 1968 on interstate traffic in guns on homicide rates and found that handgun homicide rates actually grew faster in New York City and Boston (the restrictive cities) than the average trend for fifty-seven control cities. Such time-series designs are indispensable and widely used in criminological and criminal justice research because the subjects often require analysis of trends or long-range effects rather than short-term outcomes.

Counterbalanced Designs

Counterbalanced designs are intended to manage or control the problem of multiple-treatment inference, in which X_1 refers to one treatment, X_2 a second, X_3 a third, and X_4 a fourth. By using four groups that are equivalent and by exposing each group to all four treatments and observing after each combination of treatments, it becomes possible to isolate the treatment(s), combination of treatments, or sequence of treatments that produces the outcome. For example, in four precincts where the treatments were X_1—foot patrol, X_2—media campaign, X_3—police blazers, and X_4—unarmed police, perhaps the desired outcome is obtained only by X_3—police blazers and X_1—foot patrol in that sequence of introduction only. Such a design, although complex, is the only way to uncover such a relationship.

E	X_1O	X_2O	X_3O	X_4O
E	X_2O	X_3O	X_4O	X_1O
E	X_3O	X_4O	X_1O	X_2O
E	X_4O	X_1O	X_2O	X_3O

There are many other variations of the experimental model. Familiarity with the notation and basic designs we have discussed enables one to conceptualize these other designs as offshoots of the basic ones.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

SOME OTHER CRIMINAL JUSTICE EXAMPLES OF VARIATIONS OF THE EXPERIMENTAL MODEL

The Provo and Silverlake Experiments

Empey and Erickson (1972) and Empey and Lubeck (1971) employed the classic experimental design to assess the effect of experimental community-based treatment programs in Provo, Utah, and Silverlake, Los Angeles. The subjects were all juveniles who were either serious and/or repeat offenders. All were ordinarily candidates for reformatories who were sentenced by a judge to probation or incarceration. All were male. None were seriously retarded, addicted to drugs, or had a history of serious assaultive violence.

Although both the Provo and the Silverlake experiments are relatively complex in theoretical design and methodological execution, a brief review of two elements of the analysis will serve our purposes. One element in community resistance to community-based treatment programs has been the fear of crime perpetrated by those undergoing treatment. The Provo researchers, and later the Silverlake study, compared randomly assigned experimental and control groups from two disposition conditions. Our discussion is concerned primarily with the Provo Study, although the Silverlake experience is also briefly discussed.

Provo Research Design

Probation experimental group	*E*	*O*	*X*	*O*
Probation control group	*E*	*O*		*O*
Incarceration experimental group	*E*	*O*	*X*	*O*
Incarceration control group	*E*	*O*		*O*

Equivalence of groups was obtained by randomly assigning those given probation to the Provo program or regular probation and similarly assigning those scheduled for incarceration to either the Provo treatment or actual incarceration. With the incarceration group, randomization broke down because the population was too small and the matched control group was selected using a state training school.

In evaluating arrest rates as an indicator of crime committed, the probation controls had a rate twice as high as the probation experimentals. Surprisingly, the arrest rate for "incarcerated" experimentals was almost as low as that for incarcerated controls. The latter were either on furloughs or were escapees who, in addition, tended to commit more serious crimes. Similar findings were obtained in the Silverlake experiment, which added the variables of urban setting and race (20 percent Hispanic or African American, versus 100 percent Caucasian in the Provo study).

In analysis of yet another dependent variable, postprogram arrest rates, by means of a longitudinal follow-up four years after release, the experimental probation types were not greatly superior to regular probation control types (a 71 percent reduction in crime compared with the rate four years prior to the experiment, versus 66 percent for the regular probation group). The "incarcerated" experimental group showed a 49 percent reduction, whereas the incarcerated control group exhibited only a 25 percent reduction. Maturation or age did not affect these differences when controlled statistically; that is, these differences operated independently of age. Among some methodological problems introduced in the Provo experiment was a breakdown in equivalence, because the judge sentenced too few to the reformatory.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

Additionally, a possible Hawthorne effect was present, as suggested by the fact that the success rate for the control group given regular probation was higher than that for the same group before the experiment began. The probation officers' knowledge of the study may have had this impact (Hood and Sparks, 1971, p. 209).

Despite some problems, given the higher costs of incarceration, the Provo/Silverlake findings were a major demonstration of the utility of the community-based corrections movement. Exhibit 3.3 reports on evaluations of shock incarceration.

EXHIBIT 3.3

Evaluations of Shock Incarceration

With a burgeoning prison population in the 1980s, intermediate sanctions such as electronic monitoring, intensive probation, and shock incarceration became popular, cost-effective alternatives to overcrowding prisons. They turned out not to fit the role of a popular panacea. Often increased recidivism took place due to an increased "net of control" (Morris and Tonry, 1990) or increased technical violations due to greater surveillance during community supervision (Petersilia and Turner, 1990). While boot camp programs (shock incarceration) vary in content, most involve offenders participating in military-type training and variations of physical exercise, hard physical labor, ventilation therapy, substance abuse therapy, and prerelease education. Doris Mackenzie and associates have examined a number of boot camp programs.

Examining a Louisiana boot camp (shock incarceration) program, Mackenzie and Shaw (1993) looked at shock incarceration releasees after two years of community supervision and compared these with similar offenders who had been given probation or parole. Four groups of offenders were compared: shock releasees, probationers, parolees, and shock dropouts. These were contrasted with respect to technical violations (terms of supervision offenses), new crime arrests, and new crime convictions. The shock graduates had higher rates of technical violations and revocations than the probationers and parolees, lower rates of new convictions, and, in some analyses, lower rates of arrests and revocations for new crimes. There were no differences between shock alumnae and shock dropouts in the Louisiana study (ibid.). This study was an example of the utilization of a quasiexperimental design with comparison groups in which there was no random assignment of subjects. The shock experimentals were compared with offenders who had been eligible for shock incarceration but had received other treatments instead. Possible invalidities in such a design include selection bias (in that offenders were not randomly assigned) and the fact that shock offenders were more carefully scrutinized on release, which most likely accounted for the greater number of technical violations (ibid., p. 483).

Between 1982 and 1992, the number of shock programs had increased to forty-one programs. In an evaluation of eight shock incarceration programs in Florida, Georgia, Illinois, Louisiana, New York, Oklahoma, South Carolina, and Texas, Mackenzie and Souryal (1994) found recidivism rates similar to comparable offenders who did not go through boot camps. Where shock offenders had lower rates, this may have been due to selection bias (specially selected offenders for the program). It appears that boot camp experience itself does not reduce recidivism (Travis, 1994). Successful programs were followed by six-month intensive supervision in the community. Participants in boot camps gave higher ratings to their experience and felt safer.

Sources: Mackenzie, Doris L., and James W. Shaw. "The Impact of Shock Incarceration on Technical Violations and New Criminal Activities." *Justice Quarterly* 10 (September 1993): 462–488; Mackenzie, Doris L., and Claire Souryal. *Multisite Evaluation of Shock Incarceration*. Rockville, M.D.: National Criminal Justice Reference Service, NCJ #150062, 1994; Travis, Jeremy. "Researchers Evaluate Eight Shock Incarceration Programs." National Institute of Justice Update, October 1994.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

The Kansas City Preventive Patrol Experiment

Yet another variation of the classic experimental design was employed in the *Kansas City preventive patrol study* (Kelling et al., 1974).

Experimental Group I (Reactive Patrols)	E	O_1	X_1	O_2
Experimental Group II (Proactive Patrols)	E	O_1	X_2	O_2
Control Group (Usual Patrols)	E	O_1		O_2

A fifteen-beat area of the city was divided into three matched five-beat groups. The first group was reactive in patrol procedure; that is, officers responded only to calls for service and did not deploy preventive patrols. The second group was proactive, and increased preventive patrols up to three times the normal levels. Usual deployment, or preventive patrols at their normal levels, was assigned to the control group.

Outcome variables analyzed prior to the treatments, O_1, included reported crime and victimization surveys of citizens and businesses. Posttreatment outcomes in citizen and business victimization and perception of security and reported and unreported crime showed no statistically significant differences among the three types of patrol areas studied. Despite methodological criticisms such as the location of cars withdrawn from reactive patrol, small size of beats, and small numbers used in the survey, the study suggested that police administrators have greater leeway than they supposed in patrol deployment (Davis and Knowles, 1975; Kelling and Pate, 1975; Larson, 1975; Pate et al., 1975; Chaiken, 1976). Similar replications of the Kansas City experiment essentially confirmed the same findings (Albuquerque, 1979).

In a review of the Kansas City data, critics concluded that it was not likely that randomization had been used in beat assignments (Feinberg, Singer, and Tanur, 1985). Kelling, the principal director of the project, admitted that the police selected the beats on the basis of the department's needs (Fagan, 1990, p. 110). Research by Sykes (1984) further illustrates the need to measure different outcomes of increased enforcement efforts. In examining saturation patrol as a deterrent to drunk driving, he found that it did deter some types of deviant behavior but was not a "panacea."

The Minneapolis Domestic Violence Experiment

A 1977 Police Foundation (1977, p. iv) study in Kansas City found that in the two years preceding a case of domestic assault or domestic homicide, the police had been at the address of the incident five or more times in half of the cases. This suggested that the police had an opportunity to attempt to head off domestic violence. Beginning in the late 1960s, the police had been encouraged to train their officers and utilize counseling and family crisis intervention strategies in domestic dispute cases. By the 1980s, concern for the rights of female victims, possible lawsuits against police for failure to make arrests where subsequent violence occurred, and a more conservative punitive-policy orientation led to a questioning of this policy.

In 1983, the *Minneapolis Domestic Violence Experiment* (Sherman and Berk, 1984a, 1984b; Sherman, 1985) was undertaken to attempt to provide evidence as to the most effective strategy. The research design is similar to the Kansas City Preventive Patrol Experiment—that is, it contains three groups (two with different treatments and one control). In Minneapolis, police officers volunteered to give up their discretion in handling simple (misdemeanor) domestic assaults and take whatever action was dictated by a random system: instructions written on a card and drawn from an envelope at the scene. Three different instructions were given: (1) *arrest* the suspect; (2) *separate* or remove the suspect from the scene for eight hours; or (3) *advise* and mediate.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

Minneapolis Domestic Violence Experiment

Arrest	E	O_1	X_1	O_2
Separate	E	O_1	X_2	O_2
Mediate*	E	O_1		O_2

After the police intervened, researchers attempted to interview the victims every two weeks for the next six months, as well as monitor police records to check if there were any subsequent assaults. The results appeared to be dramatic: 37 percent of the "advised" subjects and 33 percent of the "separated" subjects had recidivated (committed new assaults within six months); however, only 19 percent of the "arrested" subjects were repeaters. This reduction was accomplished even though arrest usually entailed only a night in jail.

The Minneapolis experiment has been both the most widely accepted and the most influential policy experiment of recent years; no other policy experiment has had quite the same impact on criminal justice policy. It's effect may be explained in part by the conservative tenor of the times, which emphasizes a law enforcement orientation for solving social problems; it may also be explained by the very aggressive dissemination of the study's findings (Binder and Meeker, 1991). Buzawa and Buzawa (1991) note that even though the research project was a modest pilot study with acknowledged limitations, extraordinary efforts to publicize the results in the national media resulted in premature police policy changes. In the first published replication of the study in Omaha, Nebraska, the researchers Dunford, Hiuzinga, and Elliott (1989, 1990) found that arrest alone did not have any greater impact than mediation or separation.

Binder and Meeker (1991) have provided the most thorough critique of the Minneapolis study. They cited several objections:

- The areas chosen for study were two Minneapolis precincts with the worst domestic violence rates.
- Officer participation in the study was not only voluntary, but poor. By the end of the study, about 28 percent of the cases were being processed by only three officers.
- Approximately 60 percent of the Minneapolis sample of victims and suspects were unemployed, whereas a similar study by Ray (1982) in the New York City area found that most were employed.
- Nearly 60 percent of the Minneapolis sample had previously been arrested, and only one-third had had husband–wife relationships. By comparison, in Ray's study (1982) only 10 percent had previous records, with two-thirds being married couples.
- The comparison treatments of separation or mediation by police officers without special training were not realistic comparison points.
- Other problems with the statistical analysis (Binder and Meeker, 1988) of prison crowding and of internal factors such as "officer interest in victim's story" raised further questions regarding the study's broad conclusions.

Further replications will tell us if we have too quickly embraced an "arrest panacea" for handling domestic disputes. Results from six replications seem to suggest that arrest does not work more effectively in deterring domestic assault.

The point of all of this is to suggest that research is an ongoing process and one in which replication is essential; panaceas or simple solutions based on one study are suspect.

* Mediation could be considered a third treatment or X_3.

HAMILTON, STEVEN 1428BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

If you refused to be intimidated by the researchese in this chapter and learned your *X*'s and *O*'s and how research design is a powerful tool for controlling rival causal factors, you are now conversant in the language and actually have gotten through the worst part. In the following chapters, we attempt to make you fluent in this language.

THE EXPERIMENT AS A DATA-GATHERING STRATEGY

We have thus far viewed experiments primarily from the standpoint of research design; however, the experiment is also a data-gathering strategy. Through its three key features of assuring equivalence of groups, pre- and posttests, and experimental and control groups, the experiment is a powerful strategy for research.

As a research design strategy, the experiment consists of blueprints outlining the conduct of the study. Saying that the experiment is the benchmark of comparison for other designs suggests that by using the *X* and *O* notation scheme, we can depict a basic model of a research study and also the potential strengths and weaknesses of such a design or research plan. The experiment is also a tool for data gathering, a strategy for obtaining and analyzing data. As a data-gathering strategy, the experiment has many variations that are defined by the setting. These variations range from laboratory experiments to field experiments, the former having greatest control over experimental conditions (thus high internal validity); but because of the very controlled atmosphere, problems may exist in terms of artificiality or external validity. Field experiments have fewer internal controls but greater external validity. In discussing relative advantages and disadvantages of the experiment, it is difficult to distinguish whether critics are talking about a design or data-gathering strategy or whether they are critiquing laboratory and/or field experiments or both.

Advantages of Experiments

Advantages of experiments

The **advantages of experiments** are many. They offer the best control for factors that tend to affect the internal validity of studies. The researcher is able before the fact, by the very design of the study, to control for many of the rival causal factors that tend to invalidate findings. The experimenter can control for the effects of many variables by including or excluding them from the study design.

A second advantage of experiments is that they are *relatively quick* and *inexpensive*. In contrast to many of the other data-gathering strategies that we will examine in Chapters 4–8, an experiment generally produces the required data necessary for analysis rather quickly. Depending on the scope of the study and required staff, facilities, and equipment, the experiment may represent a bargain compared with the expense of surveys or field strategies.

Another advantage of the experiment is its *manageability*, because the researcher is able to call the shots by controlling the stimulus, the environment, the treatment time, and even the degree of subject exposure. The conditions and conduct of experiments are often so rigorously defined that they lend themselves to replication by which the design and methodology can be repeated by other investigators. This is a major advantage over some field studies and surveys where it may be more difficult to repeat all of the ingredients.

Experimental strategies can be applied to natural settings in which the researcher has the best of both worlds, rigorous control and a more natural setting. If conditions can be viewed as realistic by subjects, *experiments may be the only way of studying certain complex behaviors*. Additionally, "natural experiments," which "occur as part of a natural process, where neither the setting nor the randomization process are controlled" (Fagan, 1990, p. 13), may present themselves without any

HAMILTON, STEVEN 4238BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

intervention by the researcher. For example, a new treatment program might be implemented, but it might be applied to only half of the subjects due to funding shortages.

Disadvantages of Experiments

Disadvantages of experiments

Despite the many advantages of experiments as a data-gathering strategy, there are potential disadvantages a researcher should take into account in whether experiments are the preferred strategy. The major **disadvantage of experiments** is their *artificiality*. In essence, the very controls imposed by the researcher to control for rival causal factors internal to an experiment may create artificial conditions that impede the ability to extrapolate to larger populations which are subject to natural conditions. In controlling for extraneous conditions, one may literally be creating a mere shadow of the former entity. This problem is more severe with laboratory experiments than field experiments.

In a typical critique of the contrived nature of some experiments, Field showed how many laboratory simulations of jury decision making using college students as jurors may be in error. Using randomly selected students and nonstudents in juror roles, Field (1978) found students to be significantly more lenient in their sentencing. Thus, experimental results have no assumed built-in validity. Some of the scorn with which some experimental researchers view data obtained by other social science and criminal justice researchers using field research methods is misplaced.

Other major problems relate to the general difficulty of doing experimental research in terms of obtaining human subjects or situations/conditions in which one can properly manipulate the variables to be investigated. Major ethical issues can be raised by using experimental research. Luskin points out the difficulty of implementing experimental designs in court research. Court personnel may decline to experiment with new procedures or be unable to manipulate key variables (Luskin, 1978). Hackler suggests that in evaluations of delinquency prevention programs, traditional experimental–control group procedures are nearly impossible, create unnecessary stress for program staff, and may produce hostility toward the researcher. After-the-fact statistical analysis is in this instance viewed as far less obtrusive and more useful than precontrolled studies (Hackler, 1978). Most judges are unwilling to permit treatment decisions to be governed by pure random selection. In fact, one professor of criminal law indicated that such assignment could constitute a violation of the right to due process (Glaser, 1976, p. 775), although others point out that most randomized experiments are ethical and legal (Erez, 1986). *Experimenter effects* may also occur in experiments in which those conducting the research actually selectively observe that they wish to see or unconsciously give cues to the subjects as to the desired behavior or attitudes expected. Experiments provide an excellent method for controlling for factors regarding internal validity, but they are often weak with respect to external validity.

Summary

Assumptions of causality rest at the basis of scientific investigation. Three essential steps are necessary to attempt to resolve the causality question: demonstration of relationship, specification of time order, and control for, or exclusion of, rival causal variables. The experimental model is one of the most powerful means of controlling for rival causal factors before the fact through the design of research. Rival causal factors may be treated as factors affecting internal validity and those affecting external validity. Although the former are errors introduced because of flaws within the study, the latter are factors that impinge on the generalizability of the study to larger populations. Factors affecting internal validity

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

include history, maturation, testing, instrumentation, statistical regression, experimental mortality, and interaction effects such as selection–maturation interaction. Those impacting on external validity include testing effects, selection bias, reactivity, and multiple-treatment interference. Related invalidating factors include the Hawthorne effect, the placebo effect, the halo effect, and post hoc error.

The classic experimental design is the benchmark or point of departure for all other research designs. That is, in a sense, all forms of research can be viewed as a variation of the experimental model. The three basic components of the classic experiment are pretest and posttest, experimental and control groups, and equivalence. In addition, familiarization with the notation of experimental designs is a useful heuristic device for breaking down the essentials of a research design. Classic experimental designs provide for the most rigorous before-the-fact control over factors of internal validity, and different variations enable control for rival causal factors. Illustration of the various designs with examples from criminal justice research provides the reader some familiarity with applications of these designs from the criminology and criminal justice literature.

After this review and examination of examples of the experimental model and its variations, Chapter 4 will explore the relative advantages and disadvantages of alternatives to the experimental model in criminal justice and criminology.

Experimental methods of gathering data have distinct *advantages* such as rigid control over rival factors within the experiment, the relative quick and inexpensive manner in which readily quantifiable data can be gathered, and overall manageability from the standpoint of the researcher. The *disadvantages* of experiments often outweigh their advantages, particularly in dealing with criminal justice subject matter. Major shortcomings of the experimental method include artificiality, which may hinder its generalizability to wider populations, and difficulty in applying the approach to human subjects and situations in criminal justice.

Examples of the Kansas City gun experiment, child abuse victims and violence, and shock incarceration illustrated various research designs in this chapter.

Key Concepts

Steps for Resolving the Causality Problem *67*
Rival causal factors *67*
Spurious relationship *67*
Internal validity *68*
External validity *68*
History *68*
Maturation *69*
Testing *69*
Instrumentation *70*
Statistical regression *70*
Selection bias *70*
Experimental mortality *71*
Selection–maturation interaction *71*
Testing effects *71*
Reactivity *72*
Multiple-treatment interference *72*
Hawthorne effect *72*
Halo effect *73*
Self-fulfilling prophecy *73*
Post hoc error *73*
Placebo effect *73*
Double-blind experiment *74*
Research designs *74*
Classic experimental design *74*
Equivalence *75*
Randomization *75*
Matching *75*
Pretest–Posttest *75*
Experimental group *75*
Control group *75*
Dualistic fallacy *83*
Time-series designs *83*
Advantages of experiments *92*
Disadvantages of experiments *93*

Review Questions

1. How does research design control for rival causal factors? Describe, for example, how the classic experimental design controls for history and maturation.
2. Find a recent journal article that employs an experimental, preexperimental, or quasi-experimental design. Name, describe, and illustrate the design, and discuss any rival causal factors controlled for in this design.

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.

3. Why are time-series designs particularly useful in criminal justice studies?
4. Design a hypothetical study, and discuss how your design controls for many rival causal factors.
5. Discuss the Kansas City gun experiment. What type of research design was employed, and what were the major findings of the project?

Useful Web Sites

Uniform Crime Reports *www.fbi.gov*
Sourcebook of Criminal Justice Statistics *www.albany.edu/sourcebook/*
National Incident-Based Reporting System (NIBRS) *www.ojp.usdoj.gov/bjs/nibrs.htm*
Bureau of Justice Statistics www.ojp.usdoj.gov/bjs/
CJ Ed (Researching Criminal Justice) *www.cjed.com/rschers.htm*
Jastnet (Justice Technology Information Network) *www.nlect.org/links/statlinks.html*

HAMILTON, STEVEN 4228BU

Research Methods in Criminal Justice and Criminology, 8E by Frank E. Hagan. Published by Prentice Hall.
Copyright © 2010 by Pearson Education.