10

Quasi-Experimental Designs

LEARNING OBJECTIVES

1. To learn how quasi-experimental designs differ from true experimental designs.
2. To learn the basic characteristics of each of the quasi-experimental designs.
3. To understand how rival hypotheses are ruled out in each of the quasi-experimental designs.

One of the most significant changes occurring in the life of many Americans during the last several decades of the twentieth century was the advent of the personal computer. It is estimated that by 1998 approximately 40 percent of all U.S. households owned a personal computer (Kraut, Patterson, Lundmark, Kiesler, Mukopadhyay, & Scherlis, 1998). Ever more significant is that about a third of these homes had access to the Internet, and this percentage continues to grow. The Internet is a marvelous technological development providing easy access to information and communication in a way unparalleled by anything we have known in the past. People use the Internet to obtain information previously inaccessible, to increase technological skills, and to conduct commercial transactions all in the comfort of their home. The Internet is also used for social purposes, to communicate and socialize with colleagues, friends, and family through electronic mail, and to join social groups through distribution lists, newsgroups, and chat rooms. There is some evidence (Kraut, Mukhopadhyay, Szczypula, Kiesler, & Scherlis, 1998) that the Internet is used primarily for communication with others. This has led some researchers to ask questions about the behavioral and psychological effect of the Internet. These questions seem to be particularly important because of numerous reports of individuals becoming addicted to the Internet and this addiction leading to divorce, child neglect, job termination, debt, flunking out of school, and legal trouble.

Young (1996), for example, cites the case of a homemaker who initially spent a few hours per week scanning a variety of chat rooms. During the next three months she continued to increase the time spent to converse or "chat" online with other individuals, peaking at fifty to sixty hours per week. Frequently she spent more than the intended two hours on-line with some sessions lasting up to fourteen hours. This obsession with participating in "chat" rooms eventually resulted in her reducing her involvement with her family, eliminating social activities, and ceasing to perform routine chores such as cooking, cleaning, and grocery shopping. The end result was that she became estranged from her two daughters and separated from her husband within one year of the purchase of her home computer.

The repeated occurrence of stories such as the one just mentioned and research (Kraut et al., 1998) demonstrating the negative social impact of extensive use of the Internet has led some individuals, such as Harvard psychologist Maressa Hecht Orzack (Potera, 1998) to develop a treatment for this problem. Dr. Orzack's approach is to treat Internet addiction like binge eating, where the client is taught to set limits, balance activities, and schedule time. However, this treatment has not, as yet, been empirically validated or demonstrated to be effective through sound empirically based experimental studies. It is appropriate, therefore, to ask the question Is this treatment effective? On the basis of the material you have learned in previous chapters, you should realize that this research question could be answered using a good experimental design. A sample of individuals meeting specified criteria for excessive use of the Internet could be randomly assigned to a control group that did not receive Orzack's treatment or an experimental group that did receive the treatment. Following treatment, the two groups could be compared to determine if the experimental group gained more control over their Internet addiction than did the control group.

The problem with using such an approach is that many treatment programs will not allow a researcher to come in and determine randomly whether a person can or cannot receive treatment. Rather, the treatment program states that their mission is to treat individuals, such as those with an addiction to the Internet, and they will accept anyone who requests treatment: It would be unethical to do otherwise. This is one of the primary difficulties encountered in moving out of the laboratory and into the real world. Outside the laboratory setting, it is more difficult to use control techniques, and therefore harder to control for the influence of extraneous variables. But in such cases investigators need not throw up their hands and abandon the research. Rather, they must turn to the use of quasi-experimental designs—designs that enable researchers to investigate problems that preclude the use of procedures required by a true experimental design.

This chapter will present a variety of quasi-experimental designs and will discuss the way in which the influence of rival hypotheses must be considered when these designs are used.

Introduction

Quasi-experimental design A research design in which an experimental procedure is applied but all extraneous variables are not controlled

A **quasi-experimental design** is an experimental design that does not meet all the requirements necessary for controlling the influence of extraneous variables. In most instances the requirement that is not met is that of random assignment of participants to groups. For example, several years ago the head of the probation department in a large metropolitan area approached a group of investigators, of which I was a part, and asked us to design a study to investigate the validity of the hypothesis that the food eaten by juvenile delinquents is causally related to their delinquent behavior. This administrator said that we could use the juveniles who had been committed to one of the detention facilities as our research participants. Since these youngsters were required to spend all of their time at this detention facility and the food available to them was prepared there, this was an ideal setting in which to test the nutrition–behavior hypothesis. Once we started designing the experiment, however, we encountered some of the constraints that investigators may find when moving out of the laboratory and into the real world. We were told that we could not randomly assign the juveniles into experimental and control groups; they all had to be treated in the same manner. Consequently, we realized at the outset that it would be impossible to conduct a true experiment and that we had to settle for a design that would not provide maximum assurance that the experimental and control groups were equated. In other words, we had to settle for a quasi-experimental design.

You may ask whether it is possible to draw causal inferences from studies based on a quasi-experimental design, since such a design does not rule out the influence of all rival hypotheses. Many causal inferences are made without using the experimental framework; they are made by rendering other rival interpretations implausible. If a friend of yours unknowingly stepped in front of an oncoming car and was pronounced dead after being hit by the car, you would probably attribute her death to the moving vehicle. Your friend might have died as a result of numerous other causes (a heart attack, for example), but such alternative explanations are not accepted because they are not plausible. In like manner, the causal interpretations arrived at from quasi-experimental analysis are those that are consistent with the data in situations where rival interpretations have been shown to be implausible. Of course, the identification of what is and is not plausible is not always as apparent as this illustration suggests. If it were, we would not need to conduct the experiment. I am simply demonstrating the type of procedure that must be used within the framework of quasi-experimental designs.

STUDY QUESTION 10.1

How does a quasi-experimental design differ from a true experimental design, and how are rival hypotheses ruled out in quasi-experimental designs?

Nonequivalent Control Group Design

Nonequivalent control group design A quasi-experimental design in which the results obtained from nonequivalent experimental and control groups are compared

A number of designs have been identified by Cook and Campbell (1979) as being **nonequivalent control group designs.** This kind of design includes both an experimental and a control group, but participants are not randomly assigned. The fact that participants in the control and experimental groups are not equivalent on all variables may affect the dependent variable. These uncontrolled variables operate as rival hypotheses to explain the outcome of the experiment, making these designs quasi-experimental designs. But where a better design cannot be used, some form of a nonequivalent control group design is frequently recommended. The basic scheme, depicted in Figure 10.1, consists of giving an experimental group and a control group first a

FIGURE 10.1
Nonequivalent control group design.

	Preresponse measure	*Treatment*	*Postresponse measure*	*Difference*	
Experimental group	Y_1	X	Y_2	$Y_1 - Y_2$	Compare
Control group	Y_1		Y_2	$Y_1 - Y_2$	

(Adapted from *Experimental and Quasi-Experimental Designs for Research* by D. T. Campbell and J. C. Stanley, 1963. Chicago: Rand McNally and Company. Copyright 1963, American Educational Research Association, Washington, D.C.)

pretest and then a posttest (after the treatment condition is administered to the experimental group). The pre- to posttest difference scores of the two groups are then compared to determine if significant differences exist. The design appears identical to the before–after experimental design. However, there is one basic difference that makes one a *true* experimental design and the other a *quasi*-experimental design. In the before–after design, participants are randomly assigned to the experimental and control groups, whereas in the nonequivalent control group design they are not. The absence of random assignment is what makes a design quasi-experimental.

Consider the study conducted by Becker, Rabinowitz, and Seligman (1980), which was concerned with the impact of the billing procedure on energy consumption. Because of the large energy bills resulting from increased energy costs, a number of utility companies have given their customers the option of using an "equal monthly payment plan." This scheme requires the utility company to bill the resident for one-twelfth of the yearly utility cost each month, as opposed to billing for the actual amount consumed. Although such a plan apparently produces a great deal of customer satisfaction, it runs the risk of increasing energy use, because there is no direct connection between energy used and the size of the monthly bill. The study conducted by Becker et al. was designed to determine if the equal monthly payment plan actually led to an increased use of energy.

In conducting such a study, we would ideally assign participants randomly to either the equal monthly payment plan or the conventional payment plan (in which energy is paid for as it is consumed). For a variety of reasons, however, the utility companies contacted would not allow such random assignment, so the investigators had to formulate two groups without randomization. This meant that a quasi-experimental design had to be used, and Becker et al. selected the nonequivalent control group type. Figure 10.2 shows that the design of the Becker et al. study consisted of pretesting both groups on consumption of electricity prior to the implementation of the equal payment plan. Following this pretesting (which occurred during the summer months), the treatment plan was implemented for the experimental group, and consumption of electricity was measured for both groups during the following summer.

In formulating the experimental and control groups, Becker et al. did not have the opportunity to assign participants randomly, although they were

FIGURE 10.2
The design of the Becker, Rabinowitz, and Seligman (1980) study.

	Pretest response	*Treatment conditions*	*Posttest response*
Experimental group	Magnitude of electricity consumed	Equal monthly payment plan	Magnitude of electricity consumed
Control group	Magnitude of electricity consumed	Conventional payment plan	Magnitude of electricity consumed

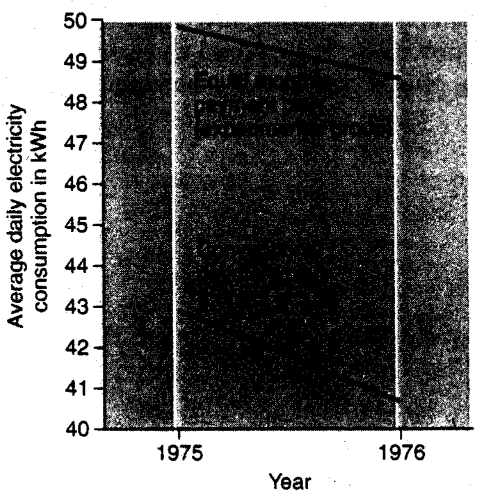
aware of the need to have equated groups. Consequently, they devised a system that seemed to match the participants on the variables that would influence electrical consumption. One of the companies whose customers were used in the study maintained records in such a way that it was possible to identify next-door neighbors. The investigators reasoned that next-door neighbors would be more likely to have similar-sized homes and to be more similar on other variables that may affect electrical consumption than would a random sample of individuals not on the equal monthly payment plan. Therefore, the control group consisted of next-door neighbors of those individuals who were on the equal monthly payment plan.

The results of this study for one company are depicted in Figure 10.3, which shows that a difference in electrical consumption existed between the two groups at pretesting time. However, the change in consumption between pretesting and posttesting was about the same for both groups. The question now becomes one of interpreting these results. The difference in pretest scores suggests that the two groups were not equivalent at the beginning of the experiment, in which case variables other than the experimental condition may have produced the obtained results. For example, those selecting the equal monthly payment plan used more electricity at the outset and, therefore, may have differed from the control group in a variety of ways.

Cook and Campbell (1979) have pointed out that the rival hypotheses in such a situation tend to be directly related to the results obtained from the experiment. These researchers identified several different experimental outcomes that could occur from the use of a nonequivalent control group design. They then listed the rival hypotheses that could also explain the obtained results. We will first take a look at these outcomes and the rival hypotheses that threaten them and then attempt to relate them to the Becker et al. study.

FIGURE 10.3
Average daily electricity consumption for two payment plans.

(Based on data from "Evaluating the Impact of Utility Company Billing Plans on Residential Energy Consumption" by L. J. Becker, V. C. Rabinowitz, and C. Seligman, 1980, *Evaluation and Program Planning, 3,* pp. 159–164.)



STUDY QUESTION 10.2 | **Diagram the nonequivalent control group design, and explain why it is a quasi-experimental design.**

Outcomes with Rival Hypotheses

Increasing treatment effect I outcome
An outcome in which the experimental and the control groups differ at pretesting and only the experimental group's scores change from pre- to posttesting

Increasing Treatment Effect I Outcome In the **increasing treatment effect I outcome,** illustrated in Figure 10.4, the control group scores reveal no change from pretest to posttest, but the experimental group starts at a higher level and shows a significant positive change. Such an outcome appears to suggest that the experimental treatment was effective. However, this outcome could also have occurred as a result of a selection– maturation effect or a local history effect.

Selection–maturation effect
The result of selecting one of two groups in such a way that its participants develop faster than those in the other group

A **selection–maturation effect** refers to the fact that one of the two groups of participants was selected in such a way that its participants were growing or developing faster than the participants in the other group. One group may progress faster because its members are more intelligent or capable than those in the other group. In the Becker et al. study, an increasing treatment effect I outcome would have been indicated if the experimental group had consumed more electricity during pretesting and had continued to increase consumption between pre- and posttesting, while control group consumption remained stable. Such an increase could have been caused by the type of payment plan used by the experimental group, but it could also have been caused by the fact that the salary level of this group was increasing, so that these participants were less concerned with electrical costs. If this were the case, then the posttest increase could be accounted for by the fact that the selection procedure happened to place in the experimental group individuals whose salary levels were increasing more rapidly.

In an attempt to eliminate the potential biasing of this type of selection–maturation effect, many investigators try to match research participants. This procedure is supposed to equate participants on the matched variables not only at the time of matching but also during the remainder of

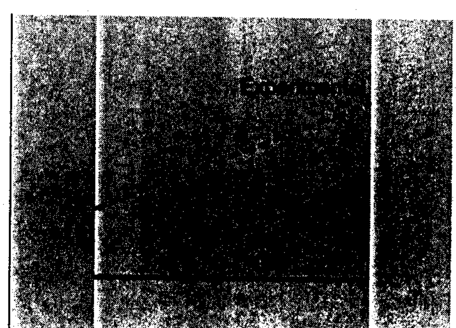


FIGURE 10.4
Increasing treatment effect I outcome.

(From "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings" by T. D. Cook and D. T. Campbell, in *Handbook of Industrial and Organizational Psychology,* edited by M. D. Dunnette. Copyright © Rand McNally College Publishing Company, 1976.)

the study. If matching is conducted during the pretest, then experimental and control participants should not differ on the dependent variable measure. If they do not, then it is assumed that they are equated. This equality is supposed to persist over time, so any difference observed during a posttest is attributed to the experimental treatment effect. However, evidence (Campbell and Boruch, 1975; Campbell and Erlebacher, 1970) has revealed that such an assumption could be erroneous because of a statistical regression phenomenon that may occur within the two groups of participants. This regression phenomenon increases the difference between the two matched groups upon posttesting, apart from any experimental treatment effect. Such a difference could be misinterpreted as being due to a treatment effect or a failure to find a treatment effect, depending on which of the matched groups operated as the experimental group and which operated as the control group.

Assume that we are conducting a study designed to investigate the influence of a Head Start program on children's subsequent school performance. We consider the attitudes of the mothers to be important, so we decide to match on this variable to eliminate its influence. Assume further that the attitude scores obtained from mothers of Head Start children and of non–Head Start children are distributed in the manner shown in Table 10.1.

TABLE 10.1
Hypothetical Attitude Scores

Head Start Participants	Head Start Mothers' Pretest Attitudes	Mothers' Posttest Attitudes	Non–Head Start Participants	Non–Head Start Mothers' Pretest Attitudes	Mothers' Posttest Attitudes
S_1	5		S_{16}	25	28
S_2	7		S_{17}	27	30
S_3	9		S_{18}	29	32
S_4	11		S_{19}	31	34
S_5	13		S_{20}	33	36
S_6	15		S_{21}	35	
S_7	17		S_{22}	37	
S_8	19		S_{23}	39	
S_9	21		S_{24}	41	
S_{10}	23		S_{25}	43	
S_{11}	25	22	S_{26}	45	
S_{12}	27	24	S_{27}	47	
S_{13}	29	26	S_{28}	49	
S_{14}	31	28	S_{29}	51	
S_{15}	33	30	S_{30}	53	

Matched Participants

From this table, it is readily apparent that most of the Head Start mothers have lower attitude scores than do the non–Head Start mothers. Therefore, matching involves selecting for the experimental group those Head Start mothers with the highest attitude scores and for the control group those non–Head Start mothers with the lowest attitude scores. In other words, we include only participants with extreme score—the participants most susceptible to the statistical regression phenomenon. This would not be a serious factor if the distributions of scores of the two groups were the same, but they are not. Statistical regression dictates that the Head Start mothers' scores, upon posttesting, will decline and regress toward the mean of their group and that the control participants' scores will regress or increase toward their group's mean (also illustrated in Table 10.1). Such a regression phenomenon could indicate that the experimental treatment is detrimental when it actually may not have any effect. If the treatment does have a positive effect, this regression effect might lead us to underestimate it.

Another way of attempting to equate participants by eliminating the selection–maturation bias artifact is to use a variety of statistical regression techniques, such as analysis of covariance and partial correlation. Campbell and Erlebacher (1970) and Campbell and Boruch (1975) have pointed out the fallacy of such an approach, but a discussion of this fallacy is beyond the scope of this text. Suffice it to say that these researchers and others (Cronbach and Furby, 1970; Lord, 1969) have found that such statistical adjustments cannot equate nonequivalent groups unless there is no error in the dependent measures given to the research participants.

Local history effect
The result of an extraneous event's influencing either the experimental or the control group, but not both groups

A second rival explanation of the increasing treatment effect I outcome is a **local history effect** (Cook and Campbell, 1975). A general history effect, discussed in Chapter 7, is controlled in the nonequivalent control group design by inclusion of a control group. However, the design is still susceptible to a local history effect, in which some event affects either the experimental or the control group but not both. A local history effect could have operated in the Becker et al. study—if the participants in the control group had purchased additional insulation for their homes, the control group would have decreased consumption of electricity not because of the type of payment plan but because of the additional insulation. Such a variable would represent a rival hypothesis for any difference observed between the control and the experimental groups.

Increasing treatment and control groups outcome
An outcome in which the experimental and the control groups differ at pretesting and both increase from pre- to posttesting, but the experimental group increases at a faster rate

Increasing Treatment and Control Groups Outcome In the **increasing treatment and control groups outcome,** both the control group and the experimental group show an increment in the dependent variable from pre- to posttesting, as is depicted in Figure 10.5. The difference between the increased growth rates could be the result of an actual treatment effect, but it could also be due to a type of selection–maturation interaction. Figure 10.5 indicates that participants in both groups are increasing in performance. Note, however, that at the time of pretesting, the treatment group scored higher on the dependent variable. This could mean that the participants in

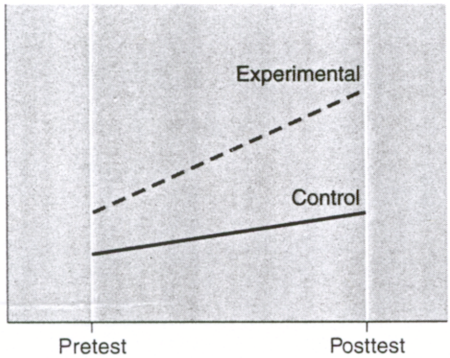


FIGURE 10.5
Increasing treatment and control groups outcome.
(From "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings" by T. D. Cook and D. T. Campbell, in *Handbook of Industrial and Organizational Psychology*, edited by M. D. Dunnette. Copyright © Rand McNally Publishing Company, 1976.)

the experimental treatment group were just naturally increasing faster on the dependent variable than were the control participants. The greater difference between groups of participants at posttesting might simply reflect the fact that the experimental participants continued to increase faster on the dependent variable than did the control participants. For example, assume that the dependent variable consisted of a measure of problem-solving ability and that the participants were six years old at the time of pretesting and eight years old at the time of posttesting. Also assume that the experimental participants were brighter and therefore increasing in problem-solving ability more rapidly than were the control participants. If this were the case, then we would expect the two groups of participants to differ somewhat at pretest time. However, participants would not stop increasing in problem-solving ability at age six, and thus an even greater difference would exist at posttest time, independent of any treatment effect. Where such a differential growth pattern occurs, we may interpret a greater posttest difference as being the result of a treatment effect when it is really an artifact of a selection–maturation interaction.

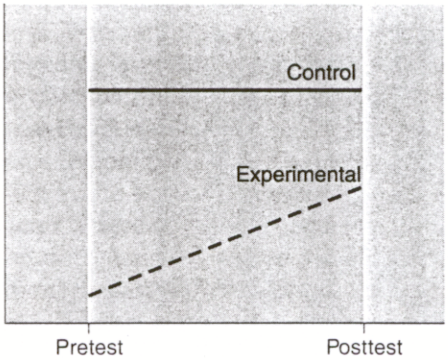
Evidence of the existence of a selection–maturation interaction can be seen by looking at the variability of the participants' scores at pretest and posttest time. Random error dictates that the variability of the scores should be the same on both occasions. A growth factor dictates that the scores should increase in terms of variability, however. Thus an increase in the variability of the scores for the experimental and control groups from pretest to posttest suggests the possibility of the existence of a selection–maturation interaction.

Increasing treatment effect II outcome
An outcome in which the control group performs better than the experimental group at pretesting but only the experimental group improves from pre- to posttesting

Increasing Treatment Effect II Outcome The **increasing treatment effect II outcome,** depicted in Figure 10.6, is an outcome in which the control group and experimental treatment group differ rather extensively at pretest time. However, the experimental group improves over time, presumably because of the experimental treatment, so that the posttest difference is decreased. Such an outcome would be desired when the experimental group was a disadvantaged group and the experimental treatment was designed to

FIGURE 10.6
Increasing treatment effect II outcome.

(From "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings" by T. D. Cook and D. T. Campbell, in *Handbook of Industrial and Organizational Psychology*, edited by M. D. Dunnette. Copyright © Rand McNally College Publishing Company, 1976.)



overcome the disadvantage. For example, Head Start was initiated to overcome the environmental deprivation experienced by many children in the United States and bring the performance of these disadvantaged individuals up to that of nondisadvantaged children. If a study were conducted to compare the pretest and posttest performances of a group of control individuals (who had not experienced the environmental deprivation) with those of a group of environmentally disadvantaged children who had received the Head Start experimental treatment, we would hope to find the type of effect illustrated in Figure 10.6. However, before we can interpret the increase in performance of the experimental treatment group as being the result of the Head Start experience, several rival hypotheses must be ruled out. The first is a local history effect that affects only one of the two groups of participants. The second and more likely rival hypothesis is a statistical regression effect—a likely source of confounding, since the participants in the experimental treatment group are typically selected because of their unusually poor performance or low scores. Consequently, the regression artifact would predict that the scores of this group should increase during posttesting. Statistical regression could, therefore, produce the outcome depicted in Figure 10.6, an outcome that the unwary investigator would interpret as a treatment effect. Therefore, designs that involve administering an experimental treatment to a disadvantaged group should provide a check for the possibility of such a regression artifact.

One indicator of the existence of a regression artifact is the instability of the deprived group's scores in the absence of the experimental treatment. If the deprived group's scores stay consistently low over time, this suggests that the low scores represent the true standing of the individuals. In such cases, a pretest-to-posttest increment would probably represent a true experimental effect, or at least an effect not confounded by the influence of a regression artifact.

Crossover effect
An outcome in which the control group performs better at pretesting but the experimental group performs better at posttesting

Crossover Effect Figure 10.7 depicts the **crossover effect,** an experimental outcome in which the treatment group scores significantly lower than the control group at pretest time but significantly higher at posttest

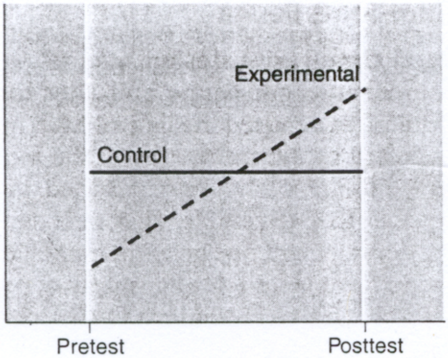


FIGURE 10.7
Crossover effect.
(From "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings" by T. D. Cook and D. T. Campbell, in *Handbook of Industrial and Organizational Psychology*, edited by M. D. Dunnette. Copyright © Rand McNally College Publishing Company, 1976.)

time. This outcome represents the typical interaction effect and is much more readily interpreted than the others discussed because it renders many of the potential rival hypotheses implausible. Statistical regression can be ruled out because it is highly unlikely that the experimental treatment group's lower pretest scores would regress enough to become significantly higher than those of the control group on posttesting. Second, a selection–maturation effect is improbable because it is typically the higher scoring pretest participants who gain faster. The outcome depicted in Figure 10.7 shows that the participants scoring lower on the pretest increased their scores more rapidly than did the control group, which scored higher on the pretest. This is the opposite of what a selection–maturation outcome would suggest.

STUDY QUESTION 10.3

Identify and discuss the rival hypotheses that could explain the different outcomes that could occur in a nonexperimental control group design. Why is the cross-over effect not readily explained by rival hypotheses?

Time-Series Design

In such research areas as psychotherapy and education, it is very difficult to find an equivalent group of research participants to serve as a control group. Is the one-group before–after design (discussed in Chapter 9) the only available design in such cases? Is there no means of eliminating some of the rival hypotheses that arise from this design? Fortunately, there is a means for eliminating some of these hypotheses, but to do so one must think of mechanisms other than using a control group. "Control is achieved by a network of complementary control strategies, not solely by control-group designs" (Gottman, McFall, and Barnett, 1969, p. 299). These complementary strategies are detailed in the following section.

Interrupted Time-Series Design

Interrupted time-series design
A quasi-experimental design in which a treatment effect is assessed by comparing the pattern of pre- and posttest scores of one group of research participants

The **interrupted time-series design** requires the investigator to take a series of measurements both before and after the introduction of some treatment condition, as depicted in Figure 10.8. The result of the treatment condition is indicated by a discontinuity in the recorded series of response measurements. Consider the study conducted by Lawler and Hackman (1969) in which they tried to identify the benefit derived from employee participation in the development and implementation of an employee incentive plan. Prior research had investigated a variety of payment plans and found that a given plan (say, a bonus plan) may be successful in one instance and not in another, indicating that the success of pay incentive plans is a function of factors other than just the plan itself. Lawler and Hackman hypothesized that a particular pay incentive plan would be more effective if the employees participated in its development, as opposed to having a plan dictated by management. To assess the validity of this theory, Lawler and Hackman had three work groups meet and develop a bonus incentive plan for reducing absenteeism. Absenteeism rates for these work groups were measured before and after the incentive plan was developed. The rates were then converted to a percentage of the number of scheduled hours that the employees actually worked. The average percentage of scheduled hours actually worked for all participants appears in Figure 10.9. From this figure, you can see that there was a rise in this average percentage and that this rise persisted over the sixteen weeks during which data were collected. All this is a visual interpretation, however. Now it is necessary to ask two questions. First, did a significant change occur following the introduction of the treatment condition? Second, can the observed change be attributed to the treatment condition?

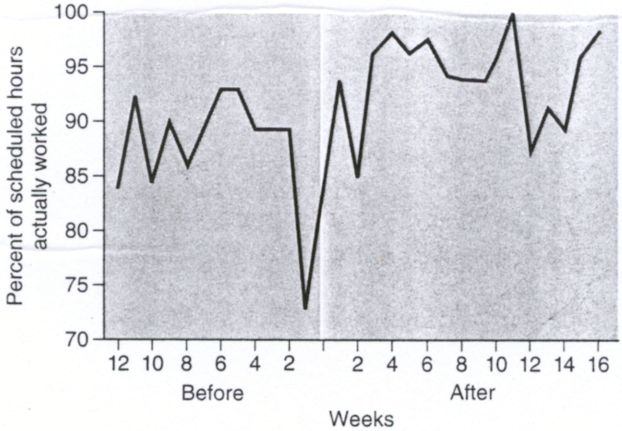
The answer to the first question naturally involves tests of significance, since, as Gottman, McFall, and Barnett (1969, p. 301) have stated, "The data resulting from the best of experimental designs is of little value unless subsequent analyses permit the investigator to test the extent to which obtained differences exceed chance fluctuations." However, before presenting the specific tests of significance, I want to follow the orientation set forth by Campbell and Stanley (1963) and Caporaso and Ross (1973) and discuss the possible outcome patterns for time series that would reflect a significant change resulting from an experimental alteration. Let us first take a look at the data that would have been obtained from Lawler and Hackman's 1969 study and a study conducted by Vernon, Bedford, and Wyatt (1924) if they

FIGURE 10.8
Interrupted time-series design.

Preresponse measure	*Treatment*	*Postresponse measure*
Y_1 Y_2 Y_3 Y_4 Y_5	X	Y_6 Y_7 Y_8 Y_9 Y_{10}

(Adapted from *Experimental and Quasi-Experimental Designs for Research* by D. T. Campbell and J. C. Stanley, 1963, American Educational Research Association, Washington, D.C.)

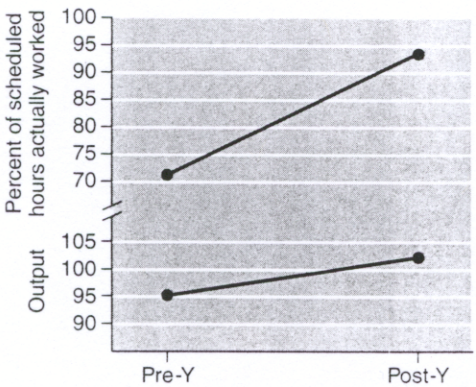
FIGURE 10.9
Mean attendance of the participative groups for the twelve weeks before the incentive plan and the sixteen weeks after the plan. (Attendance is expressed in terms of the percentage of hours scheduled to be worked that were actually worked.)



(From "Impact of Employee Participation in the Development of Pay Incentive Plans: A Field Experiment" by E. E. Lawler and J. R. Hackman, 1969, *Journal of Applied Psychology, 53,* pp. 467–471. Copyright 1969 by the American Psychological Association. Reprinted by permission of the author.)

had used only a one-group before–after design. The Vernon et al. study was concerned with investigating the influence of introducing a rest period on the productivity of various kinds of factory workers. These data are presented in Figure 10.10. Note that in *both* studies beautiful data seem to support the hypothesis that the experimental treatment condition produced a beneficial effect. Remember, however, that the one-group before–after design does not include a comparison group, so the increase in performance could have been due to many variables other than the experimental treatment condition.

FIGURE 10.10
A one-group before–after representation of a portion of the Vernon et al. data and Lawler and Hackman data.



One means of eliminating some of the sources of rival hypotheses is to take a number of pre- and postmeasurements or to conduct an interrupted time-series analysis. When this kind of study is undertaken, we find the data depicted in Figure 10.9 for Lawler and Hackman's study and the data depicted in Figure 10.11 for the Vernon et al. study. The data suggest that the treatment condition investigated by Lawler and Hackman was influential but that the treatment condition investigated by Vernon et al. was not. The pattern of responses obtained by Vernon et al. seems to represent a chance fluctuation rather than a real change in performance.

Visual inspection of a pattern of behavior can be very helpful in determining whether an experimental treatment had a real effect. Caporaso and Ross (1973) have presented a number of additional possible patterns of behavior, shown in Figure 10.12, that could be obtained from time-series data. Note that the first three patterns reveal no treatment effect but merely represent a continuation of a previously established pattern of behavior. Lines D, E, F, and G represent *true* changes (I am assuming that they would be statistically significant) in behavior, although line D represents only a temporary shift.

Now let us return to our original question of whether a significant change in behavior followed the introduction of the treatment condition. Such a determination involves tests of significance. The most widely used and, I believe, the most appropriate statistical test is the Bayesian moving average model (Box and Jenkins, 1970; Box and Tiao, 1965; Glass, Tiao, and Maguire, 1971; Glass, Willson, and Gottman, 1975). Basically, this method consists of determining whether the pattern of postresponse measures differs from the pattern of preresponse measures. To make such an assessment

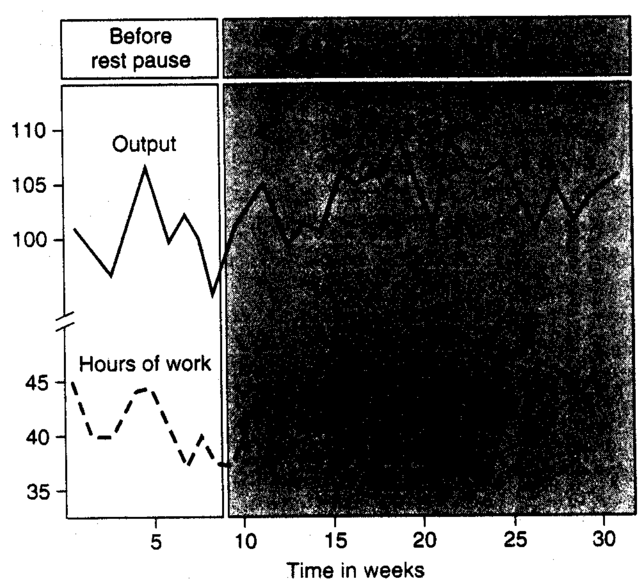


FIGURE 10.11
Effect of a ten-minute rest pause on worker productivity.
(Reprinted from *Two Studies of Rest Pauses in Industry* by H. M. Vernon, T. Bedford, and S. Wyatt, 1924. Medical Research Council, Industrial Fatigue Research Board No. 25. London: His Majesty's Stationery Office.)

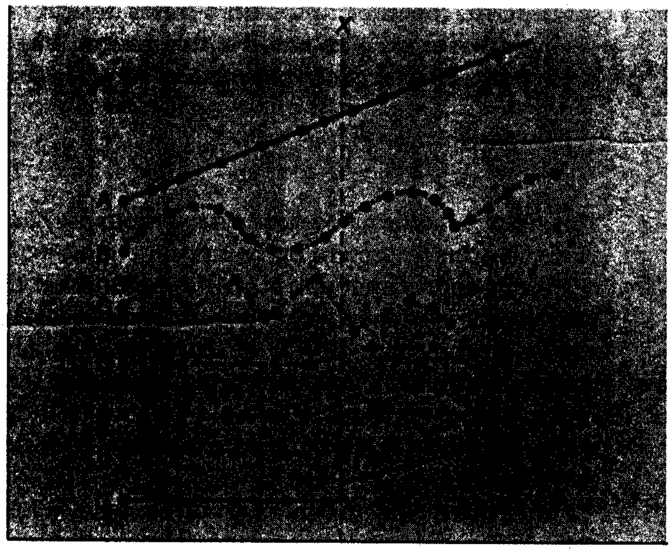


FIGURE 10.12
Possible pattern of behavior of a time-series variable.
(From *Quasi-Experimental Approaches: Testing Theory and Evaluating Policy*, edited by J. A. Caporaso and L. L. Ross, Jr. ©1993 by Northwestern University Press. Reprinted by permission.)

using the moving average model requires many data points. Glass, Willson, and Gottman (1975) recommend that at least fifty data points be obtained. This relatively large number of data points can typically be obtained when conducting experiments using animals. However, it frequently cannot be obtained when conducting research with humans. When fewer data points are obtained, the probability of concluding that a treatment is effective when it really is not, a type I error, is increased. This difficulty has resulted in a limited use of the moving average model in analyzing time-series data. Fortunately, Tryon (1982) and Crosbie (1993) have developed statistical procedures that are effective with as few as ten data points. A valid statistical analysis can, therefore, be conducted on almost any study using a time-series analysis.

Lawler and Hackman's analysis of their data revealed a significant difference between the patterns of pre- and postresponse measures. This led them to conclude that a nonrandom change occurred following the introduction of the incentive plan. This brings us to the second question: whether this significant change can be attributed to the employees' participation in the incentive plan. The primary source of weakness in the interrupted time-series design is its failure to control for the effects of history. Considering Lawler and Hackman's study, assume that at about the same time the treatment condition was introduced, some extraneous event occurred that could also have led to an increase in the number of hours worked. Such an extraneous event serves as a rival hypothesis for the significant nonrandom change. The investigator must consider all the other events taking place at about the same time as the experimental event and determine whether they might be rival hypotheses. Actually, Lawler and Hackman included several other control groups in their study to rule out such effects.

STUDY QUESTION 10.4

Describe the interrupted time-series design, and explain how rival hypotheses are eliminated in this design. What is the primary rival hypothesis that cannot be controlled when using this design?

Summary

This chapter has deviated considerably from the orientation taken in the previous chapters by presenting a number of quasi-experimental designs, which represent approximations of true experimental designs in the sense that they use the experimental mode of analysis in investigating areas that do not allow for complete control of extraneous variables. Quasi-experimental designs are the best type of design available for use in some field studies in which one wants to make causal inferences. Of the different quasi-experimental designs presented, one is a before–after design and one is a time-series design.

The nonequivalent control group design, a before–after design, is the one most frequently used. It is exactly like the before–after true experimental design except that participants are not randomly assigned to the experimental and control groups, which means that we do not have the necessary assurance that the two groups of participants are equated. We could attempt to equate participants on the important variables using matching techniques. However, this still does not assure us that the participants are totally equated, and it may produce a statistical regression effect. Nonetheless, the design is useful in that it may control for effects such as history and maturation.

The time-series design attempts to eliminate rival hypotheses without the use of a control group. In the interrupted time-series design, a series of measurements is taken on the dependent variable both before and after the introduction of some experimental treatment condition. The effect of that condition is then determined by examining the magnitude of the discontinuity produced by the condition in the series of recorded responses. The primary source of error in this design is the possible history effect.

Key Terms and Concepts

- Quasi-experimental design
- Nonequivalent control group design
- Increasing treatment effect I outcome
- Selection-maturation effect
- Local history effect
- Increasing treatment and control groups outcome
- Increasing treatment effect II outcome
- Crossover effect
- Interrupted time-series design