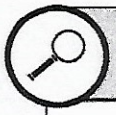

RESEARCH IN PRACTICE 10.1 Program Evaluation: Matching and Randomization in a Community-Level Alcohol Prevention Experiment

The evidence–based practice (EBP) approach promotes the idea that knowledge derived from controlled experiments provides the best evidence about which human service practices and policies work. The Center for Substance Abuse Treatment (CSAT) is part of the Substance Abuse and Mental Health Services Administration (SAMHSA). SAMHSA promotes EBP through the National Registry of Evidence-Based Programs and Practices (NREPP), which is a searchable database of interventions for the prevention and treatment of mental and substance use disorders (http://www.nrepp.samhsa.gov). Developed to help people, agencies, and organizations implement programs and practices in their communities, the service is one more way that researchers and practitioners are becoming more closely linked.

NREPP includes an easy to use interface for locating relevant research. For example, we located the experiment discussed below by simply selecting "Substance Abuse Prevention" as an "Area of Interest" in the "Advanced Search Mode". The search yielded 96 studies, including Communities Mobilizing for Change on Alcohol (CMCA) (Wagenaar, Murray, & Toomey, 2000). The CMCA program is a community organizing program designed to reduce teens' access to alcohol.

To evaluate the program, the research team used school district criteria to insure a sample of comparable participants and to minimize influence from other factors besides the independent variable such as impact of other community programs and interaction between intervention and control communities. Because communities and not individual students were the target of this study, the researchers applied randomization at the community level. Twenty-four eligible school districts were identified, and 15 agreed to participate. After all baseline data were collected, the 15 participating districts were matched on population size and presence of a college. From within matched sets, seven communities were randomly selected to receive the community organizing intervention, and the remaining eight were assigned to the control group.

The intervention itself consisted of a community organizer working with local public officials, enforcement agencies, alcohol merchants and merchant associations, the media, schools, and other community institutions to change community policies to reduce youth access to alcohol. The target of the intervention was the entire community rather than individual young people. The organizing effort resulted in institutional policy changes such as alcohol merchant

single test administration, however, a certain number of people probably will score lower than their normal level. If the whole group repeats the exam—even without any intervening experimental manipulation—they can be expected to perform better; thus, the average score of the group will improve.

Selection. *Selection* is a threat to internal validity when the kinds of people who are selected for one experimental condition differ from those who are selected for other conditions. The threat of selection derives from improperly constituted experimental and control groups. Recall the importance of random assignment or matching to equalize these groups. As noted, improper composition of these groups can make findings based on comparisons between them meaningless.

Experimental Attrition. The threat of *experimental attrition* occurs when there is a differential dropout of subjects from the experimental and control groups. Especially in experiments that extend over long periods of time, some people fail to complete the experiment. People die, move away, become incapacitated, or simply quit. If there is a notable difference in attrition rates between the experimental and control groups, those groups may not be equivalent at the end of the experiment, even though they were at the beginning. An example of the attrition threat can be found in a study concerning the effects of sex composition on the patterns of interaction in consciousness-raising counseling groups (Carlock & Martin, 1977). One group was all women, and the other was both men and women. The women were randomly assigned to one group or the other. Originally, the all-

policies and practices, increases in media coverage of alcohol issues, and changes in practices of law enforcement agencies.

Several outcome measures were employed, including surveys of high school students, surveys of youth age 18 to 20, and surveys of alcohol retailers. It is well known that self-reported survey data of sensitive information is subject to bias and underreporting, so even though that data was supportive of the program, this field experiment was made stronger by the fact that the researchers also used direct observation and archival data to measure program impact. Alcohol purchase attempts were conducted at bars, restaurants, liquor stores, and convenience stores. Archival data consisted of drunk driving (DUI) arrest data and traffic crash data for a six-year baseline prior to the program and three years of intervention. The researchers hypothesized the CMCA intervention would reduce DUI and disorderly conduct arrests and traffic crashes, particularly among 18- to 20-year-olds, since alcohol consumption was reduced among this age group.

The researchers reported that the CMCA organizing process was successful in changing institutional policies to reduce youth access to alcohol and decreased alcohol consumption among 18- to 20-year-olds (Wagenaar et al., 2000). The archival data on arrests and vehicle crashes suggest that this intervention also reduced alcohol-related problems among youth. Reductions were observed in all arrest and traffic crash indicators for both age groups.

For the practitioner responsible for helping a community to reduce alcohol-related problems, this study is encouraging. Although anyone seeking to design an intervention program might locate this study simply by conducting an independent literature search, the fact that an organization such as SAMHSA provides the National Registry of Evidence-Based Programs and Practices accomplishes several things:

- Locating research studies is quick and efficient.
- The search process locates groups of studies that meet the searcher's criteria.
- The user can be confident in the quality and rigor of the search results because not only the studies themselves but the review process and the selection criteria are carefully constructed and reviewed by experts in the field.

female group contained nine members, and the male-and-female group had 16 members. Before the experiment was completed, however, two female members of the male-and-female group dropped out. When the interactional patterns in the groups were compared, substantial differences were found. The authors concluded that these differences resulted from the differing sex compositions of the groups, but, as they noted, confidence in their findings was weakened because of the dropouts from the male-and-female group. With such small groups to begin with, the loss of two members from one group can substantially affect the results.

With this lengthy list of threats to the internal validity of experiments, can we ever be confident that changes in the dependent variable are the result of the independent variable and not some extraneous variability? In fact, such confidence can be established—and threats to internal validity controlled—through the use of good experimental designs.

EXPERIMENTAL DESIGNS

There are three categories of experimental designs. **Preexperimental designs** lack both the random assignment to conditions and the control groups that are such a central part of good experimental designs. Although they sometimes are still useful, they illustrate some inherent weaknesses in terms of establishing internal validity. The better designs are called *true experimental designs* and

quasi-experimental designs. **True experimental designs** are more complex and use randomization and other techniques to control the threats to internal validity. **Quasi-experimental designs** are special designs that are used to approximate experimental control in nonexperimental settings.

Preexperimental Designs

On the surface, Design P1 in Table 10.2 might appear to be adequate. The subjects are pretested, exposed to the experimental condition, and then tested again. Any differences between the pretest and posttest measures should be the result of the experimental stimulus. As it stands, however, this design has serious weaknesses. With the exceptions of selection and experimental attrition, which are irrelevant because of the lack of a control group, Design P1 is subject to the other five threats to internal validity. If a historical event related to the dependent variable intervenes between the pretest and the posttest, its effects could be confused with those of the independent variable. Maturational changes in the subjects also could produce differences between pretest and posttest scores. If paper-and-pencil measures are used, then a shift in scores from the pretest to the posttest could occur because of testing effects. Regardless of the measurement process, instrumentation changes could produce variation in the pretest and posttest scores. Finally, if the subjects were selected because they possessed some extreme characteristic, differences between pretest and posttest scores could be the result of regression toward the mean. In all these cases, variation on the dependent variable produced by one or more of the validity threats could easily be mistaken for variation produced by the independent variable.

The other preexperimental design—the Static Group Comparison, or Design P2 in Table 10.2—involves comparing one group that experiences the experimental stimulus with another group that does not. In considering this design, it is important to recognize that the comparison group that appears to be a control group is not a control group in the true sense. The major threat to validity in this design is selection. Note that no random assignment, which would make the comparison groups comparable, is indicated. In Design P2, the group compared with the experimental group is normally an intact group picked up only for the purpose of comparison. There is no assurance of comparability between it and the experimental group. For example, we might wish to test the impact of a local assistance program by comparing a city that has the program to one that does not. Any conclusions we might reach about the effects of the program could be inaccurate because of other differences between the two cities.

TABLE 10.2 Preexperimental Designs

P1 The One Group Pretest–Posttest Design:

	pretest	treatment	posttest
Experimental Group:	O	X	O

P2 The Static Group Comparison:

	pretest	treatment	posttest
Experimental Group:		X	O
Control Group:			O

Despite their weaknesses, preexperimental designs are used when resources do not permit the development of true experimental designs. Human service practitioners especially are likely to be faced with this dilemma. Conclusions based on such designs should be regarded with the utmost caution, however, and the results viewed as suggestive at best. These designs should be avoided if at all possible. If they are used, however, then efforts should be made to further test the validity of the findings using one of the true experimental designs.

True Experimental Designs

The Classic Experimental Design. Diagrams of true experimental designs are presented in Table 10.3. Probably the most common true experimental design is Design T1 in Table 10.3—the Pretest–Posttest Control Group Design with Randomization, which is identical to the design on

BLE 10.3 True Experimental Designs

The Pretest-Posttest Control Group Design with domization—the "classic" experimental design:

		pretest	treatment	posttest
mental Group:	R	O	X	O
ol Group:	R	O		O

The Solomon Four-Group Design:

		pretest	treatment	posttest
mental Group 1:	R	O	X	O
ol Group 1:	R	O		O
mental Group 2:	R		X	O
ol Group 2:	R			O

The Posttest-Only Control Group Design:

		pretest	treatment	posttest
mental Group:	R		X	O
ol Group:	R			O

The Multiple Experimental Group with One Control p Design:

		pretest	treatment	posttest
mental Group 1:	R	O	X_1	O
mental Group 2:	R	O	X_2	O
mental Group 3:	R	O	X_3	O
ol Group:	R	O		O

The Factorial Design:

		pretest	treatment	posttest
mental Group 1:	R	O	X_1Y_1	O
mental Group 2:	R	O	X_1Y_2	O
mental Group 3:	R	O	X_2Y_1	O
ntrol Group:	R	O	X_2Y_2	O

ge 259 except for randomization. This design is ed so often that it frequently is referred to by its pular name: the *classic* experimental design. In all e experimental designs, the proper test of potheses is the comparison of posttests between perimental and control groups. This design uses a e control group, including random assignment equalize the comparison groups, which elimi- es all the threats to internal validity except for rtain patterns of attrition. Because of this, we can have considerable confidence that any differences between experimental and control groups on the dependent variable result from the effect of the independent variable.

Let's take a closer look at how the classic design avoids the various threats to validity. History is removed as a rival explanation for differences between the groups on the posttest, because both groups experience the same events except for the experimental stimulus. Because the same amount of time passes for both groups, maturational effects should be equal and, therefore, not account for posttest differences. Similarly, because both groups are pretested, any testing influences on the posttest should be the same for the two groups. Instrumentation effects also are readily controlled with this design, because any unreliability in the measurement process that could cause a shift in scores from pretest to posttest should be the same for both comparison groups.

In situations where regression can occur, the classic experimental design controls it through random assignment of subjects with extreme characteristics. Thus, whatever regression does take place should be the same for both groups. Regression toward the mean, therefore, should not account for any differences between the groups on the posttest. Randomization also controls the validity threat of selection by assuring that the comparison groups are equivalent. Furthermore, the pretest results can be used as a check on precisely how similar the two groups actually are. Because the two groups are similar, attrition rates should be about the same for each group. In a lengthy experiment with a large sample, we would fully expect about the same number of subjects in each group to move away, die, or become incapacitated during the experiment. These can be assumed to be more or less random events. Attrition because of these reasons, therefore, is unlikely to create a validity problem. Attrition because of voluntary quitting, however, is another matter. The experimental stimulus may affect the rate of attrition. That is, subjects might find something about the experimental condition either more or less likable than the subjects find the control condition, so the dropout rate

could differ. If this occurs, it raises the possibility that the groups are no longer equivalent at the time of the posttest. Unfortunately, there really is no effective way of dealing with this problem. About all that researchers can do is to watch for its occurrence and, if attrition bias appears to be a problem, interpret the results cautiously.

The Solomon Design. A second true experimental design—the Solomon Four-Group Design (Design T2 in Table 10.3)—is more sophisticated than Design T1 in that it uses four different comparison groups. In comparing Designs T1 and T2, we see that the first two groups of the Solomon Design constitute Design T1, indicating that it is capable of controlling the same threats to internal validity as that design does. The major advantage of the Solomon Design is that it can tell us whether changes in the dependent variable result from some *interaction effect* between the pretest and the exposure to the experimental stimulus. Experimental Group 2 is exposed to the experimental stimulus without being pretested. If the posttest of Experimental Group 1 differs from the posttest of Experimental Group 2, this may be the result of an interaction effect of receiving both the pretest and the experimental stimulus—something that we would not find out with the classic design.

Suppose, for example, that we wanted to assess the effect on prejudice toward racial minorities (the dependent variable) of receiving positive information about a racial group (the independent variable). We pretest groups by asking them questions regarding their prejudice toward particular groups. Then, we expose them to the experimental stimulus: newspaper articles reporting on civic deeds and rescue efforts by members of those racial groups. Lower levels of prejudice in Experimental Group 1 than in Control Group 1 might be caused by the independent variable. It also could be, however, that filling out a pretest questionnaire on prejudice sensitized people in the first group to these issues and, thus, that they reacted more strongly to the experimental stimulus than they would have without such pretesting. If this is the case, Experimental Group 2 should show less change than Experimental Group 1. If the independent variable has an effect that is separate from its interaction with the pretest, Experimental Group 2 should show more change than Control Group 1. If Control Group 1 and Experimental Group 2 show no change but Experimental Group 1 does show a change, the change is produced only by the interaction of pretesting and treatment.

Thus, the Solomon Design enables us to make a more complex assessment of the causes of changes in the dependent variable. In addition, the combined effects of maturation and history can be controlled (as with Design T1) and measured. By comparing the posttest of Control Group 2 with the pretests of Experimental Group 1 and Control Group 1, these effects can be assessed. Usually, however, our concern with history and maturation effects is only in controlling their effects, not in measuring them.

Despite the superiority of the Solomon Design, it often is passed over for Design T1 because the Solomon Design requires twice as many groups, thus substantially increasing the time and cost of conducting the experiment. Not surprisingly, many researchers decide that the advantages are not worth the added cost and complexity. If a researcher desires the strongest design, however, then Design T2 is the one to choose.

The Posttest-Only Control Group Design. Sometimes, pretesting is either impractical or undesirable. For example, pretesting might sensitize subjects to the independent variable. In these cases, we can still use a true experimental design—the Posttest-Only Control Group Design (see Design T3 in Table 10.3). Despite the absence of pretests, Design T3 is an adequate true experimental design that controls threats to validity as well as the designs with pretests do. The Posttest-Only Control Group uses random assignment to conditions, which distinguishes it from the preexperimental Design P2. The only potential question of validity raised in conjunction with Design T3 is selection: The absence of pretests means that random assignment is the only assurance that the comparison groups are equivalent. Campbell and

Stanley (1963), however, argued convincingly that pretests are not essential, and that randomization reliably produces equivalent groups. Furthermore, they argued that the lack of popularity of Design T3 stems more from the tradition of pretesting in experimentation than from any major contribution to validity from using pretesting.

The Multiple Experimental Group with One Control Group Design. The preceding experimental designs are adequate for testing hypotheses when the independent variable is either present (experimental group) or absent (control group). Yet, many hypotheses involve independent variables that vary in terms of the *degree* or *amount* of something that is present. For example, *how much* treatment of a client (intensity level) is needed to produce some desired behavioral change? In a study of abusive parents, the independent variable was exposure to positive parenting sessions (Burch & Mohr, 1980). Rather than just assess the effect of the presence or absence of such sessions, however, the researchers exposed some groups to *more* sessions than others. In cases such as this, the Multiple Experimental Group with One Control Group Design (see Design T4 in Table 10.3) can be used. This is an extension of Design T1 with multiple experimental groups and one control group. The symbols X_1, X_2, and X_3 refer to differing amounts or intensities of a single independent variable or treatment. Comparing the posttests of the experimental groups enables us to determine the impact of the differing amounts of the independent variable.

The Factorial Design. The experimental designs considered thus far can assess the impact of only a single independent variable at a time on the dependent variable. We know, however, that variables can *interact* with one another, and the combined effects of two or more variables may be quite different from the effects of each variable operating separately. For example, one experimental study investigated the impact of both social class and race on stereotypes of women (Landrine, 1985). People in that study were asked to describe society's stereotype of four different women: a middle-class black, a middle-class white, a lower-class black, and a lower-class white. Landrine found that social class and gender did affect stereotyping: Lower-class people received more negative stereotyping than middle-class people, and blacks were viewed less favorably than whites. Landrine did not, however, find an interaction effect. In other words, various class/race combinations did not produce more changes in stereotyping than the effects of class and race did separately.

The experimental design used in the Landrine study is called a Factorial Design (see Design T5 in Table 10.3). Design T5 illustrates the simplest factorial design, in which two independent variables (X and Y, or race and social class) each have only two values (X_1 or X_2 and Y_1 or Y_2). In factorial designs, each possible combination of the two independent variables constitutes one of the experimental conditions. Technically, there may not be a control group, as in the example above, when there can be no "absence" of the factors involved. After all, one cannot be without race or without social class. If the variables involve the presence or absence of something, however, then a condition that looks like a control group appears in the factorial design. Factorial designs involve enough groups so all possible combinations of the independent variables can be investigated. Assessment of interaction is accomplished by the comparison of Experimental Group 1—which is exposed to both independent variables—with Experimental Groups 2 and 3. We can see whether the combined effects of the independent variables differ from their separate effects.

Factorial designs can be expanded beyond this simple example. More than two variables can be investigated, and each variable can have more than two values. As the complexity increases, however, the number of groups required increases rapidly and can become unmanageable. For example, with three independent variables, each with three values, we would need a $3 \times 3 \times 3$ factorial design, or 27 different groups. Sometimes, however, field experiments designed to assess the impact of social programs involve such complex designs.

RESEARCH IN PRACTICE 10.2 Program Evaluation: Field Experiments on Police Handling of Domestic Violence Cases *(Continued)*

addition to arrest, the police could issue citations requiring a court appearance by both victim and offender, or they could advise and separate (a combination of the mediation and separation treatments in the other studies). An additional explicit criterion addressed the concern of victim and officer safety: Cases were excluded if the victim insisted on arrest, if the assailant threatened or assaulted the police, or if the officers believed that the offender posed an imminent danger to the victim. After restoring order at the scene, all victims were provided information regarding community resources, including the Victim Assistance Program and the battered women's shelter.

Choosing the best way to operationalize key variables is an important lesson from these studies. A related issue was whether to rely on official arrest records or interview reports of victims (Maxwell, Garner, & Fagan, 2001). If repeat incidents of domestic assault are measured by rearrest of the perpetrator, repeat incidents appear to be rather small in number. If the occurrence of repeat events is measured by victims' reports to interviewers, however, repeat incidents appear to be common, which is another example of the issue of validity discussed in Chapter 5—namely, how accurately does a measurement tool measure some theoretical concept?

The original Minneapolis study, based on 314 cases, concluded that arrest significantly reduced future assaults, and the policy implication appeared to be clear-cut: Implement an aggressive, pro-arrest policy to reduce spousal assault. Based on a total sample of thousands in the six cities included in the various replications, the general conclusions now are less clear-cut and more cautious and complicated (Maxwell, Garner, & Fagan, 2001; Sherman, 1992):

- Arrest does, in general, reduce future episodes of domestic violence, but the effect is modest.
- Arrest increases domestic violence among people who have nothing to lose, especially the unemployed.
- A small but chronic portion of all violent couples produces the majority of domestic violence incidents.

The research underscores the point that social problems and societal responses are complex and unlikely to yield singular solutions.

We have covered the major types of true experimental designs in this section. Often, however, circumstances require researchers to use a variant of one of these designs. Research in Practice 10.2 illustrates such a variation that, nonetheless, retains randomization. The designs presented here also form the basis for many other more complex designs that are capable of handling such problems in experimenting as multiple independent variables and order effects.

Quasi-Experimental Designs

Many times it is impossible, whether for practical or other reasons, to meet the conditions necessary to develop true experimental designs. The most common problems include an inability to assign people randomly to conditions and the difficulty of creating a true control group with which to compare the experimental groups. Although these can be problematic for experiments in any context, they are especially acute in field experiments and practice settings. Rather than rule out experimentation in such settings, however, the researcher may be able to use a *quasi-experimental design* that approaches the level of control of true experimental designs in situations where the requirements of the latter cannot be met. Quasi-experimental designs afford considerable control, but they fall short of the true experimental designs and, therefore, should be used only when conditions do not allow the use of a true experimental design (Achen, 1986). As we will see in Chapter 11, quasi-experimental designs also form the basis of single-system designs used in clinical practice.

The Time-Series Design. One of the simplest—and most useful—of the quasi-experimental designs is the Time-Series Design (see Design Q1 in Table 10.4), which involves a series of repeated

measures, followed by the introduction of the experimental condition, and then another series of measures. The number of pretest and posttest measurements can vary, but it is unwise to use fewer than three of each. Time-series designs can be analyzed by graphing the repeated measures and inspecting the pattern that is produced. We can conclude whether the independent variable produced an effect by observing when, over the whole series of observations, changes in the dependent variable occur.

Figure 10.1 illustrates some possible outcomes of a time-series experiment. Of major interest is what occurs between O_4 and O_5, because these are the measures that immediately precede and follow the experimental stimulus. The other measures are important, however, because they provide a basis for assessing the change that occurs between O_4 and O_5. As Figure 10.1 shows, some outcomes suggest that the stimulus has had an effect, whereas others do not. In cases A, B, and C, stimulus X appears to produce an effect. In each case, the differences between measures O_4 and O_5 are greater than the differences between any other adjacent measures. Cases E and F illustrate outcomes where we can infer that changes result from something other than the stimulus, because the differences between measures O_4 and O_5 are not substantially different from those between some other adjacent measures. The results of a time-series design are not always clear-cut, as illustrated by case D. The sharp change between measures O_5 and O_6 could be a delayed effect of X, or it could be an effect of something else. More careful analysis would be required to reach a firm conclusion in that case.

In field studies involving a time-series analysis, the experimental stimulus often is not something the researcher manipulates; rather, it is something that occurs independently of the research. Thus, any natural event presumed to cause changes in people's behavior can serve as an experimental stimulus. In a study of how the criminal justice system handles violence against women, for example, the experimental stimulus was the passage of the Violence Against Women Act of 1994 (Cho & Wilke, 2005). The VAWA was intended to provide better social service interventions for women who experience intimate partner violence, and to improve the law enforcement response in such cases. The dependent variables included such measures of the impact of the VAWA as the incidence of domestic violence, rates of perpetrator arrest, and rates of reporting incidents to the police. The data was provided by the National Crime Victimization Survey of the Bureau of Justice Statistics (see p. 202). These measurements were done for each three-month period from 1992 to 2003. The researchers found a significant decrease in the rate of intimate partner violence and an increase in the arrest of perpetrators. The patterns shown in the time-series design gave them the confidence to conclude that the passage of the VAWA was at least partly responsible for these changes.

The time-series design fares quite well when evaluated on its ability to control threats to internal validity. With the exception of history, the other threats are controlled by the presence of the series of premeasures. Maturation, testing, instrumentation, statistical regression, and experimental attrition produce gradual changes that operate between all measures. As such, they cannot account for any sharp change occurring between measures O_4 and O_5. Because this design does not use a control group, selection is not a factor. Thus, history is the only potential threat to internal validity. It is always possible that some extraneous event could intervene between measures O_4 and O_5 and produce a change that could be confused with an effect of X. In many cases, there may not be an event that could plausibly produce the noted effect, and we could be quite sure it was caused by X. Nevertheless, the inability of the time-series design to control the threat of history is considered a weakness.

The Different Group Time-Series Design. The time-series design requires that we be able to measure the same group repeatedly over an extended period. Because this requires considerable cooperation from the subjects, there may be cases when the time-series design cannot be used. A design that avoids this problem, yet maintains the other

TABLE 10.4 Quasi-Experimental Designs

Q1 The Time-Series Designs:

	pretests	treatment	posttests
Experimental Group:	$O_1O_2O_3O_4$	X	$O_5O_6O_7O_8$

Q2 The Different Group Time-Series Design:

	pretests	treatment	posttests
R	O_1		
R	O_2		
R	O_3		
R	O_4		
R		X	
R			O_5
R			O_6
R			O_7
R			O_8

Q3 The Multiple Time-Series Design:

	pretests	treatment	posttests
Experimental Group:	$O_1O_2O_3O_4$	X	$O_5O_6O_7O_8$
Control Group:	$O_1O_2O_3O_4$		$O_5O_6O_7O_8$

characteristics of the time-series design, is the Different Group Time-Series Design (see Design Q2 in Table 10.4). Rather than make repeated measures on one group, this design substitutes several randomly selected groups, each of which is measured only once but at different times. Design Q2 produces the same type of data as the regular time-series design and can be analyzed using the same graphing procedure. In terms of internal validity, Design Q2 controls the same threats as the time-series design. Random sampling is relied on to equate the several comparison groups. Design Q2 also has the same weakness as the time-series design: Historical events can intervene between measures O_4 and O_5, possibly confusing the results.

Because of the many random samples that this design requires, it is limited to situations in which the cost of those samples is low. For example, if it is possible to draw the samples and collect data by telephone, then the design is ideal. This is essentially the approach that commercial pollsters use

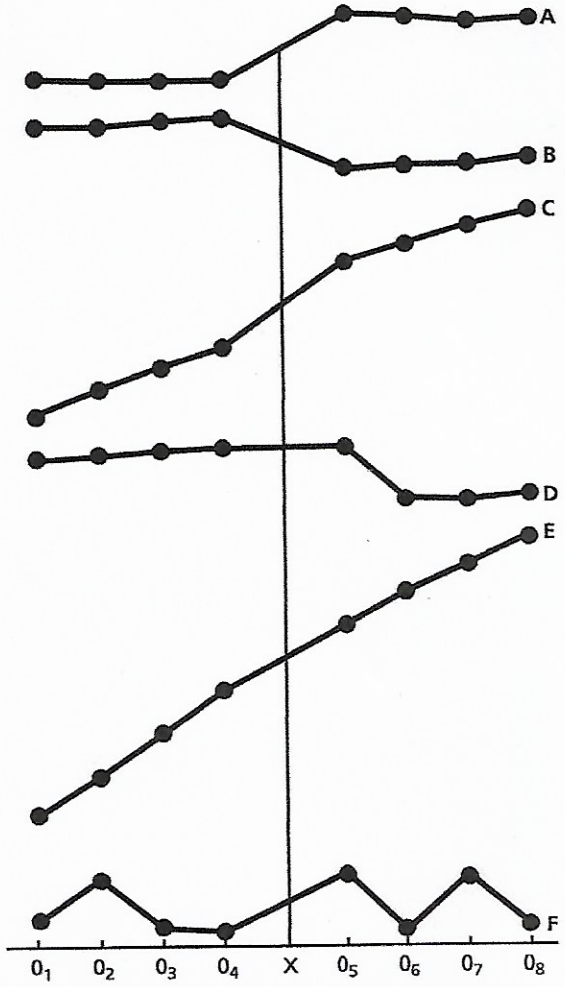

FIGURE 10.1 Possible Outcomes in a Time Series Quasi-Experimental Design

to measure public opinion by conducting weekly or monthly telephone surveys of different random samples. Although pollsters do not conduct experiments, their repeated measures allow the assessment of the impact of events on public opinion. The event in question becomes the X in the design, and the levels of opinion are compared before and after X occurred. For example, the terrorist attacks on New York and Washington in 2001 and the Iraq War in 2003 each caused the popularity of President George W. Bush to climb considerably.