CHAPTER 6

Group Experimental Designs

Do intensive, in-home, family-based interventions result in higher rates of family reunification than traditional services? Or are any observed differences due to some other factor unrelated to the interventions? Can we conclude that the intervention *caused* improved family reunification rates or are the differences due to something else? These were questions Elaine Walton and her colleagues (Lewis, Walton, & Fraser, 1995; Walton, 2001; Walton, Fraser, Lewis, Pecora, & Walton, 1993) asked in a series of studies of children in substitute care. The answers to such questions are crucial for agencies looking to implement effective services. Similarly, understanding causality is essential for understanding the roots of a social condition or examining the implications of social policy.

This chapter considers the meaning of causality, the criteria for achieving causally valid explanations, and the ways in which researchers seek to meet these criteria using group experimental research designs. By the end of this chapter, you should be able to weigh the evidence from experimental research studies about the effectiveness of social work interventions and to design group experimental research projects. In subsequent chapters, we will show how causal criteria are established in other research designs.

2 Causal Explanation

A cause is an explanation for some characteristic, attitude, or behavior of groups, individuals, or other entities (such as families, organizations, or communities). A **causal effect** means that the variation in an independent variable will be followed by variation in the dependent variable, when all other things are equal. Of course, *all* other things cannot literally be equal. We are not able to compare the same people at the same time in exactly the same circumstances except for the variation in the independent variable (King, Keohane, & Verba, 1994). However, we can design research to create conditions that are very comparable so that we can isolate the impact of the independent variable on the dependent variable. Walton et al. (1993) tried to create comparable conditions to test whether families receiving intensive, in-home, family-based services have higher rates of reunification with their children than families receiving traditional casework services; the only difference was the type of intervention received by participants.

Causal effect The finding that change in one variable leads to change in another variable, *ceteris paribus* (other things being equal).

Five criteria should be considered when deciding whether a causal connection exists. The first three criteria must be established to identify a causal effect: (1) empirical association, (2) time order, and (3) nonspuriousness. Evidence that meets the other two criteria— (4) identifying a causal mechanism and (5) specifying the context in which the effect occurs—can considerably strengthen causal explanations.

Research designs to establish these criteria require careful planning, implementation, and analysis. Many times researchers have to leave one or more of the criteria unmet and are left with some important doubts about the validity of their causal conclusions, or they may avoid even making any causal assertions.

Association

The first criterion for identifying a causal effect is an empirical (observed) **association** (sometimes a *correlation*) between the independent and dependent variables. This means that when the independent variable changes, the dependent variable also changes; if there is no association between two variables, there cannot be a causal relationship. Exhibit 6.1 displays the association that Walton and her colleagues (1993)

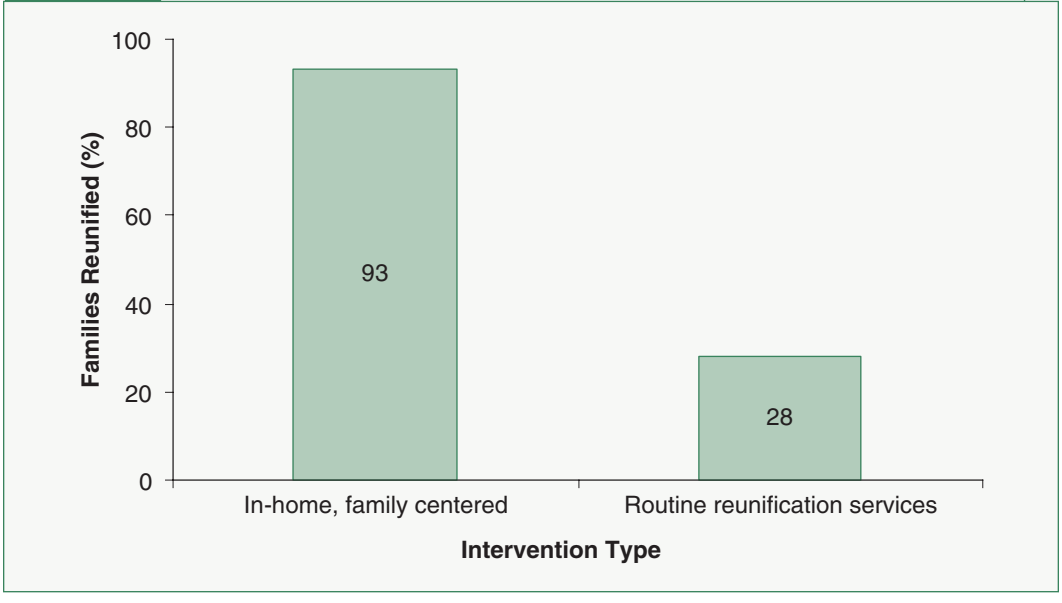
found between the type of intervention and the rate of family reunification. Family reunification after 90 days was much more common among families receiving intensive, in-home, family-based services than families receiving traditional casework services. Therefore, variation in treatment (the independent variable) is associated with family reunification (the dependent variable). An empirical association is typically demonstrated by using statistical techniques to show that the relationship between an independent and dependent variable is not due to chance.

Time Order

Association is a necessary criterion for establishing a causal effect, but it is not sufficient. We must also ensure that the change in the independent variable occurred before the change in the dependent variable. This is the criterion of **time order**. Walton's study satisfied this criterion because all participants started with the same status—families with children out of the home. Walton and her colleagues (1993) measured changes in that status, both during the 90 days of the intervention and after the 90 days were finished. It was only during and after the interventions that children could return home.

As you can imagine, we cannot be so sure about time order when we use a cross-sectional research design (see Chapter 2) to test a causal hypothesis. For example, there was evidence from cross-sectional studies that women who had given birth as teenagers had fewer years of schooling than women whose children's first birth came after the women were in their 20s (or older). What the research could not answer was whether teenage moms had fewer years of education because they dropped out of school when they had a child, or whether teenage girls who dropped out of school were more likely to become pregnant than those who stayed in school. You cannot establish a causal relationship until you establish which comes first.

Exhibit 6.1 **Association: Relationship Between Intervention Type and Family Reunification**



Source: Walton et al. (1993).

Nonspuriousness

Even when research establishes that two variables are associated and that variation in the independent variable precedes variation in the dependent variable, we cannot be sure that we have identified a causal relationship between the two variables. Before we can conclude that a causal relationship exists, we must have reason to believe that the relationship is **nonspurious**, that is, the relationship is not due to any other variable. *Spurious* means false or not genuine. A **spurious relationship** occurs when the association between two variables is due to changes in a third variable. For example, do storks bring babies? There is an empirical relationship; the more storks that appeared in certain districts in Holland, the more babies were born. But the association in Holland between the number of storks and the number of babies is spurious. The relationship is explained by a third variable, because both the number of storks and the birth rate were higher in rural districts than in urban districts. The rural or urban character of the districts (the third variable) caused variation in the other two variables. When we can rule out alternative explanations or spurious relationships, the study is said to have **internal validity**.

Mechanism

A causal **mechanism** is the process that creates the connection between variation in an independent variable and variation in the dependent variable (Cook & Campbell, 1979; Marini & Singer, 1988). In social work research, this might involve understanding the theoretical components of the intervention model, or it might be to understand the underlying process, the essential actions, that led to the desired changes. For example, Lewis et al. (1995) found that family reunification using in-home, family-based services is more effective when caseworkers have sufficient time to focus on communication skills, parenting skills, and anger management, while less effective when caseworkers devote their time to providing transportation, defusing crises, and managing conflicts.

Figuring out some aspects of the process by which the independent variable influenced the variation in the dependent variable can increase confidence in our conclusion that there was a causal effect (Costner, 1989). However, there may be many components to the causal mechanism and we cannot hope to identify them all in one study. Walton and her colleagues (Lewis et al., 1995) concluded that future research should identify the factors associated with goal achievement and success, "for goal achievement appears to be a proximal outcome measure that mediates family risk factors and more distal outcomes such as child out-of-home placement" (p. 279). The point is that identification of a mechanism through which the independent variable influences the dependent variable increases our confidence in the conclusion that a causal connection does indeed exist.

Context

No cause has its effect apart from some larger **context** involving other variables. For whom, when, and in what conditions does this effect occur? A cause is really one among a set of interrelated factors required for the effect (Hage & Meeker, 1988; Papineau, 1978). Identification of the context in which a causal effect occurs is not a criterion for a valid causal conclusion, and it is not always attempted, but it does help us to understand the causal relationship. It is important to consider when assessing the applicability of an intervention to different settings.

Walton et al. (1993) noted the possibility that contextual factors influenced their findings. The study was done in Utah and the authors suggested that Utah's unique religious and social characteristics might have influenced the findings (and limited the generalizability of the findings). Furthermore, they suggested that the decision to place a child is not just the result of the family situation but of many system-level factors that could play a role in placement decisions.

2 Threats to the Validitiy of Experimental Designs

Group experimental designs, like any research design, must be evaluated for their ability to yield valid conclusions. Remember there are three kinds of validity: (1) internal validity (nonspuriousness), (2) external validity (generalizability), and (3) measurement. In Chapter 4 we discussed measurement validity, and group experimental designs offer no special advantages or disadvantages in measurement. Therefore, we focus on internal and external validity.

Threats to Internal Validity

An experiment's ability to yield valid conclusions about internal validity is determined by the comparability of experimental and comparison groups. First, a comparison group must be created. Second, this comparison group must be so similar to the experimental group that it shows what the experimental group would be like if it had not received the experimental treatment. The following sections discuss various threats to internal validity that occur frequently in social work research (see Exhibit 6.2). These threats exemplify five major types of problems that arise in research design.

Noncomparable Groups

The problem of noncomparable groups occurs when the experimental group and the comparison group are not comparable—that is, when something interferes with the two groups being essentially the same at the start (or end) of the experiment. **Selection bias** occurs when a comparison group and experimental group are initially different and is related to the methods used to assign subjects to the groups. When subjects are assigned randomly (by chance) to treatment and comparison groups, the threat of selection bias is reduced; whereas, when subjects are assigned using other methods besides random assignment, the threat of selection bias is great. Imagine assigning highly motivated subjects to one group and less motivated subjects to a second group. The highly motivated subjects are apt to perform better than the less motivated subjects. Even if the researcher selects a comparison group that matches the treatment group on important variables (e.g., demographic characteristics), there is no guarantee that the groups are similar initially in terms of the dependent variable or some other characteristic that influences scores taken after the treatment.

Even when random assignment works as planned, the groups can become different over time because of **mortality** or differential attrition. The groups become different when subjects are more likely to drop out of one of the groups. There are different reasons why participants drop out including: (a) the study is too lengthy; (b) participants in the experimental group may become more motivated than comparison subjects to continue in the experiment; or (c) participants receiving some advantageous program benefit are more likely to stay in the experiment than participants who are not receiving program benefits.

Endogenous Change

Endogenous change occurs when natural developments in the subjects, independent of the experimental treatment, account for some or all of the observed change between a pretest and posttest. Endogenous change includes three specific threats to internal validity:

Exhibit 6.2 Summary of Threats to Internal Validity

Noncomparable groups. When characteristics of the experimental and comparison group subjects differ. These include:

- *Selection bias* Two groups are different at the start of the experiment.
- *Mortality* Group composition changes due to dropping out during the experiment.

Endogenous change. When the subjects develop or change during the experiment as part of an ongoing process independent of the experimental treatment. These include:

- *Testing* Respondents learn from the pretest: reduced anxiety from having taken the test or learning content on their own
- *Maturation* Respondents change as part of the natural process of the passing of time.
- *Statistical regression* Respondents improve or get worse because their initial scores may reflect a bad day as opposed to their true score.

External events. Events occur outside the context of the experiment but affect the participants. These include:

- *History* An event outside the experiment that participants are exposed to.
- *Secular drift.* Broader societal trends that impact on the desired outcomes of the experiment.
- *Instrumentation* Use of measures lacking reliability.

Contamination. When the comparison or control group is affected by the treatment group. This includes:

- *Compensatory rivalry* Comparison group, aware that they are denied some advantage, increase their efforts to succeed.
- *Resentful demoralization* Comparison group, aware that they are denied some advantage, reduce their efforts to succeed.
- *Diffusion of treatment* When groups interact and the nature of the treatment becomes known to the control group.

Treatment misidentification. The subjects experience something other than what the researchers believe they have experienced. This includes:

- *Compensatory equalization of treatment* Staff provide more than expected to comparison group.
- *Placebo effect* When subjects received a treatment they consider likely to be beneficial and improve because of that expectation.

- *Testing.* Taking a pretest can influence posttest scores. Participants may learn something or may be sensitized to an issue by the pretest and, as a result, respond differently the next time they are asked the same questions on the posttest. Just taking a test the first time often reduces anxiety provoked by the unknown and subjects will be more comfortable with subsequent testing.
- *Maturation.* Changes in outcome scores during experiments that involve a lengthy treatment period may be due to **maturation**. Participants may age, gain experience, or grow in knowledge, all as part of a natural maturational experience, and therefore respond differently on the posttest than on the pretest. For example, after the death of a cherished family pet, feelings of depression and sadness become less intense with the passing of time.
- *Statistical Regression.* People experience cyclical or episodic changes that result in different posttest scores, a phenomenon known as a **statistical regression**. Participants who are chosen for a study

because they received low (or high) scores on a test may show improvement in the posttest simply because some of the low scorers were having a bad day when they were assessed.

External Events

External events during the experiment (things that happen outside the experiment) could change subjects' outcome scores. One problem is **history**, or an event that subjects are exposed to during the course of the experiment or evaluation. For example, a new cook is hired at a nursing home and the food improves at the same time a researcher is testing a group intervention to improve the morale of residents.

Broader social or economic trends may also impact on the findings of a study creating a problem called **secular drift**. For example, trends in the economy may have impacted on the outcomes of welfare reform in 1996. The decline in the number of people receiving cash assistance from Aid to Families with Dependent Children began prior to 1996. It is reasonable to wonder if the Temporary Assistance for Needy Families (TANF) program produced the subsequent decline or if the reduction reflected a trend that had already begun as the economy improved.

Another possibility is **instrumentation**. When the same method of measurement is used for the pretest and posttest, the measures must be stable (demonstrate measurement reliability), otherwise the findings may reflect the instability of the measurement and not the effect of the treatment. When different methods of measurement are used, such as a paper measure for the pretest and behavioral observations for the posttest, the two methods must be equivalent (again measurement reliability); otherwise any changes might be due to the lack of equivalency.

Contamination

Contamination occurs in an experiment when the comparison group is in some way affected by, or affects, the treatment group. When comparison group members are aware that they are being denied some advantage, they may increase their efforts to compensate, creating a problem termed **compensatory rivalry** (Cook & Campbell, 1979). Comparison group members may become demoralized (called **resentful demoralization**) if they feel that they have been left out of some valuable treatment and perform worse than they would have outside the experiment. The treatment may seem, in comparison, to have had a more beneficial effect than it actually did. Both compensatory rivalry and resentful demoralization may distort the impact of the experimental treatment. Another form of contamination occurs when treatment and control (comparison) groups interact and the nature of the treatment becomes known to the control group. This problem, called **diffusion of treatment**, may result in the control group sharing in the benefits of the treatment.

Treatment Misidentification

Treatment misidentification occurs when some process that the researcher is not aware of is responsible for the apparent effect of treatment. Treatment misidentification has at least two sources:

- *Expectancies of experimental staff.* Change among experimental subjects may be due to the positive expectancies of the experimental staff who are delivering the treatment rather than the treatment itself. Even well-trained staff may convey their enthusiasm for an experimental program to the subjects in subtle ways. Such positive staff expectations create a self-fulfilling prophecy. Staff providing services to the comparison group may feel that it is unfair and, therefore, work harder or do more than they might have if there had been no experiment. This effect is called **compensatory equalization of treatment.** To counter these effects, some researchers use **double-blind procedures**, in which neither staff nor participants know whether they are in the group getting the treatment or the comparison group.

- *Placebo effect.* In social work research, a **placebo effect** may occur when participants think that something will improve (e.g., behavior) through an experimental treatment and then it does—not from the treatment but their own beliefs and expectations. Medical research indicates that the placebo effect produces positive health effects in two-thirds of patients suffering from relatively mild medical problems (Goleman, 1993).

Process analysis is a technique researchers use to avoid treatment misidentification. A researcher might periodically review audio or videotapes of the intervention to determine whether it was delivered as planned.

Generalizability

The need for generalizable findings can be thought of as the Achilles' heel of group experimental design. The design components that are essential to minimize the threats to internal validity make it more difficult to achieve sample generalizability (being able to apply the findings to some clearly defined larger population) and cross-population generalizability (generalizability across subgroups and to other populations and settings).

Sample Generalizability

Participants who can be recruited for a laboratory experiment, randomly assigned to a group, and kept under carefully controlled conditions for the study's duration are unlikely to be a representative sample of any large population of interest to social work researchers. Can they be expected to react to the experimental treatment in the same way as members of the larger population?

Researchers can take steps both before and after an experiment to increase a study's generalizability. Participants can be selected randomly from the population of interest, and thus, the researchers can achieve results generalizable to that population. Some studies of the effects of income supports on the work behavior of poor people have randomly sampled people within particular states before randomly assigning participants to experimental and comparison groups. But in most experiments, neither random selection from the population nor selection of the entire population is possible. Potential subjects must make a conscious decision to participate—probably resulting in an unrepresentative pool of volunteers.

External Validity

Researchers are often interested in determining whether treatment effects identified in an experiment demonstrate cross-population generalizability, that is, hold true for subgroups of subjects and across different populations, times, or settings. Of course, determining that a relationship between the treatment and the outcome variable holds true for certain subgroups does not establish that the relationship also holds true for these subgroups in the larger population, but it suggests that the relationship might have **external validity**.

External validity The applicability of a treatment effect (or noneffect) across subgroups within an experiment or across different populations or settings.

Evidence of an overall sample effect does not mean that the effect holds true for subgroups within the study. For example, Roger Roffman et al. (1997) examined the effectiveness of a 17-session HIV prevention group with gay and bisexual males. The researchers found that subjects in the treatment group had better outcomes than subjects in the control group. But within the treatment group, outcomes were better for exclusively gay males than for bisexual males. This study shows that an interaction effect can limit the generalizability of the findings.

Reactivity

A variant on the problem of external validity, called **reactivity**, occurs when the experimental treatment has an effect only when the particular conditions created by the experiment occur. Without the experimental conditions, there would be no effect. This is a problem as social work providers try to translate research findings into practice. The agency does not want to have to recreate the experimental conditions in order to provide an effective treatment. Reactivity takes several different forms:

- *Interaction of testing and treatment.* One such problem occurs when the treatment has an effect only if subjects have had the pretest. The pretest sensitizes the subjects to some issue, so that when they are exposed to the treatment, they react in a way they would not have reacted if they had not taken the pretest. In other words, testing and treatment interact to produce the outcome.
- *Reactive effects of experimental arrangement.* Members of the treatment group change in terms of the dependent variable because their participation in the study makes them feel special. Experimental group members could feel special simply because they are in the experiment. This is called a *Hawthorne Effect*, named after a famous productivity experiment at the Hawthorne electric plant outside Chicago. No matter what conditions researchers changed in order to improve or diminish productivity, the workers seemed to work harder simply because they were part of a special group.
- *Interaction of selection and treatment.* This effect occurs when the results are related to selection biases in who receives the treatment and who serves in the comparison group. For example, voluntary clients often do better than involuntary clients. If the treatment group consists of voluntary clients and the comparison group consists of involuntary clients, the findings of the study are likely to be influenced by the biased assignment.
- *Multiple treatment interference.* This refers to clients or subjects who have been exposed to other interventions prior to the experiment. The question of multiple treatment interference is: Was the intervention successful on its own or was it successful because of the subject's cumulative experience with other treatments or interventions? For example, chronically mentally ill individuals are likely to have had past treatment experiences, both in the community and in an institutional setting. If multiple treatment interference is a problem, the generalizability of the findings may be limited to a population having experienced a similar treatment pattern.

2 Why Experiment

Experimental research provides the most powerful design for testing causal hypotheses, because it allows us to confidently establish the first three criteria for causality: association, time order, and nonspuriousness. True experimental research designs (or randomized clinical trials) are used when a social work researcher wants to show that an intervention (independent variable) caused a change in an outcome (the dependent variable). **True experiments** have at least three features that help provide the strongest evidence about an intervention's effectiveness including:

1. Two comparison groups (in the simplest case, an experimental and a control group) to establish association

2. Random assignment to the two (or more) comparison groups to establish nonspuriousness (or internal validity)
3. Variation in the independent variable before assessment of change in the dependent variable to establish time order

To establish an association between an independent variable and a dependent variable, true experimental designs have at least two groups. One group called the **experimental group** receives some treatment or manipulation of the independent variable. The second group, the **control group**, does not receive the treatment, or as is often the case in social work research, the second group is a **comparison group,** which typically receives the traditional intervention (or treatment as usual). The experimental group and the control group (or comparison group) scores on the outcome or dependent variable are compared in order to establish an association.

A study can have more than one experimental group if the goal is to test several versions of the treatment (the independent variable) or several combinations of different treatments. For example, a researcher testing different interventions for depression might include a control group, a group receiving medication, a group receiving counseling, and a group receiving both medication and counseling. But there is still a comparison group either not receiving any treatment or receiving treatment as usual.

It is crucial that the two groups be more or less equal at the beginning of the study. **Random assignment** (or *randomization*) is used to make the experimental group and the control group similar at the beginning of the experiment (see Exhibit 6.3). You randomly sort the participants into the two groups using some chance procedure. You can do this by flipping a coin for each participant, pulling names out of a hat, using a random numbers table, or a random number generator. In any case, the subjects should not be free to choose the group nor should you (the researcher) be free to put subjects into whatever group you want.

Note that the random assignment of subjects to experimental and comparison groups is not the same as taking a random sample of individuals from some larger population (see Exhibit 6.4). In fact, random assignment does not help at all to ensure that the research subjects are representative of some larger population; instead, representativeness is the goal of random sampling. What random assignment does is ensure internal validity, not generalizability.

Why is random assignment useful for ensuring internal validity? The underlying assumption of random assignment is that if chance is used to determine who goes into a particular group, equivalent groups are created. The groups are believed not only to be more or less equal in demographic makeup and to have

Exhibit 6.3 Random Assignment to One of Two Groups

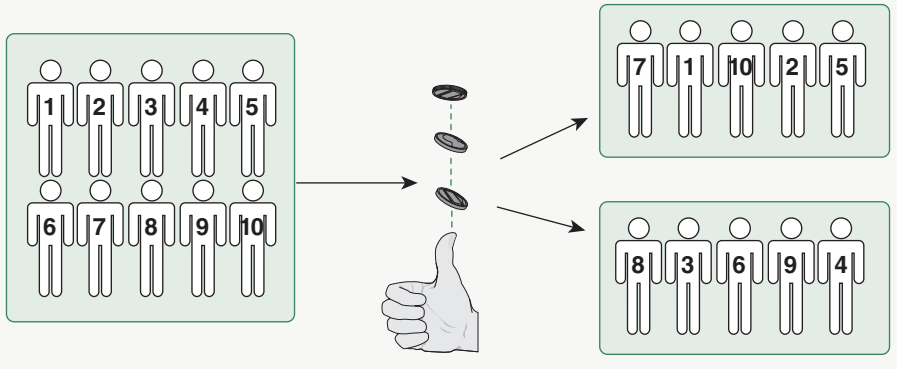
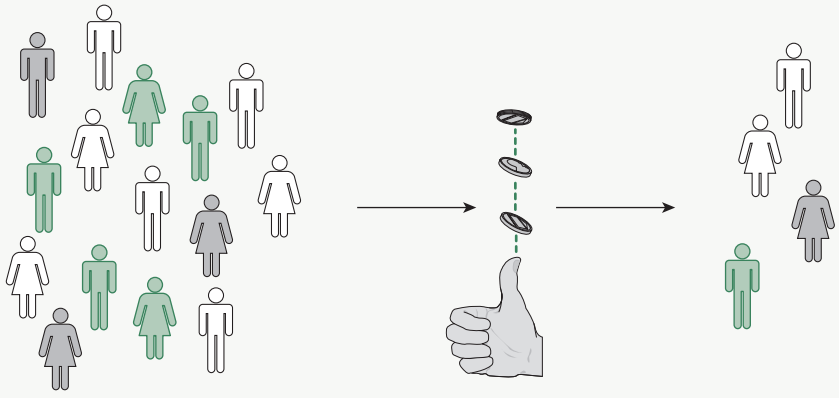
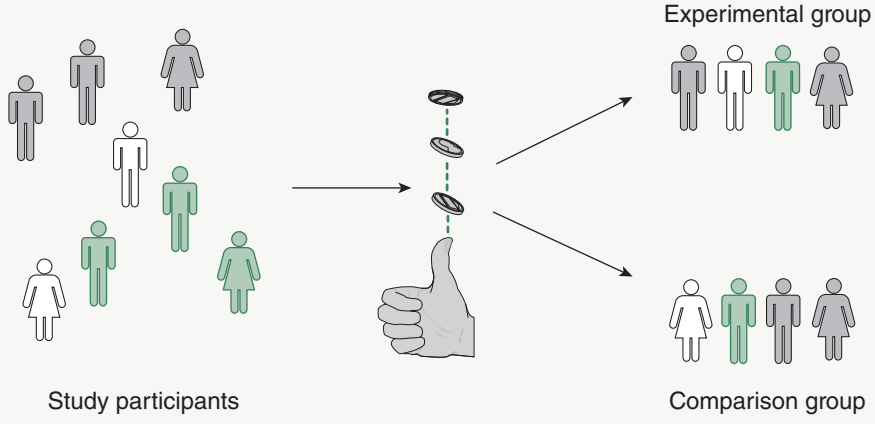


Exhibit 6.4 Random Sampling Versus Random Assignment

Random sampling (a tool for ensuring generalizability):
Individuals are randomly selected from a population to participate in a study.



Random assignment, or randomization (a tool for ensuring internal validity):
Individuals who are to participate in a study are randomly divided into an experimental group and a comparison group.



similar scores on the dependent variable (something we can check) but also to be more or less equal with regard to the impact of different possible explanations. For example, some people are highly motivated, whereas others are less motivated. Using random assignment improves the probability that highly motivated and less motivated participants are more or less equally distributed into the two groups and, therefore, motivation for change does not explain the treatment's effects.

Assigning subjects randomly to the experimental and control or comparison groups ensures that systematic bias does not affect the assignment of subjects to groups. Of course, random assignment cannot guarantee that the groups are perfectly identical at the start of the experiment. Randomization removes bias from the assignment process by relying on chance, which itself can result in some intergroup differences.

Matching is sometimes used to better equate the experimental and comparison groups. In **individual matching** individuals are matched in pairs (see Exhibit 6.5). You start by identifying important characteristics that might impact on the study, and then you match pairs of individuals with similar or identical characteristics. In a study of older adults, you might match subjects by gender, race, and age and then randomly assign each member of a pair to the experimental and control groups. This method eliminates the possibility of differences due to chance in the gender, race, and age composition of the groups.

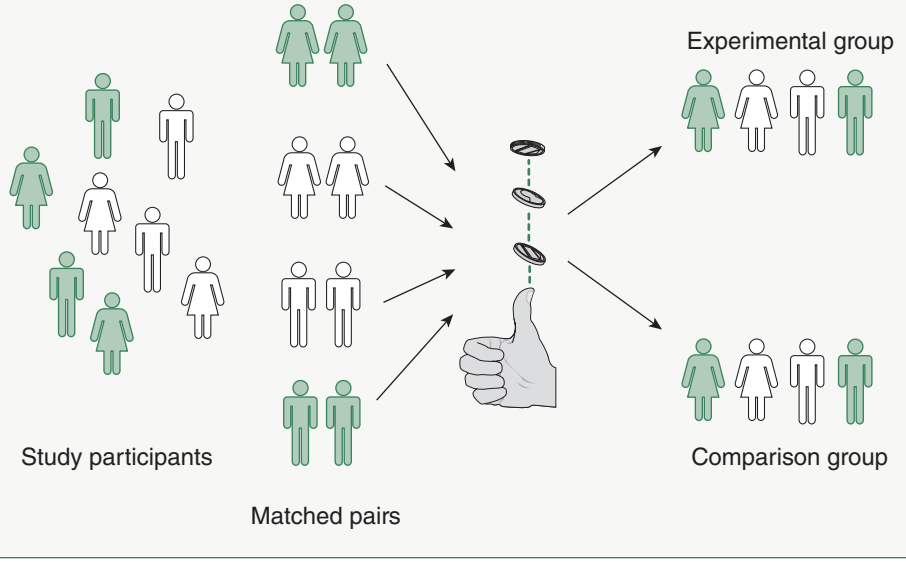
The basic problem is that, as a practical matter, individuals can be matched on only a few characteristics; unmatched differences between the experimental and comparison groups may still influence outcomes. However, matching combined with randomization can reduce the possibility of differences due to chance. A second problem occurs when one of the matched pair drops out of the study, unbalancing the groups. In this case, researchers will often exclude the findings of the individual who remained in the study.

Block matching is used when participants are grouped by their characteristics. Rather than creating pairs of older adults, the researcher might group by age and gender, so that there is a group of men between the ages of 65 and 74, a second group between the ages of 75 and 84, and a third group ages 85 and older. The same grouping by age would be done with female participants. Within each group, the members are randomly assigned into the experimental and treatment group.

Aggregate matching is matching by group. Older residents living in a high rise might be grouped by floor. The floor and not the individual residents is assigned randomly into the experimental and comparison groups. Random assignment into each group would be done by floor. Of course, this assumes that there are no systematic differences among the floors, such as apartment types.

Finally, to establish that the change in the dependent variable occurred after the intervention (time order), all true experiments have a **posttest**, that is, a measurement of the outcome in both groups after the experimental group has received the treatment. Most true experiments also have **pretests** that measure the dependent variable prior to the experimental intervention. Generally, the pretest measurement of the dependent variable is exactly the same as the posttest measurement of the dependent variable. We say *generally* because sometimes the measurement methods used for the pretest and posttest differ, although both methods must be equivalent measures of the same construct.

Exhibit 6.5 Experimental Design Combining Matching and Random Assignment



Strictly speaking, a true experiment does not require a pretest. When researchers use random assignment, the groups' initial scores on the dependent variable and all other variables are likely to be similar. Any difference in outcome between the experimental and comparison groups should be due to the intervention (or to other processes occurring during the experiment), and the likelihood of a difference just on the basis of chance can be calculated. The advantage of having a pretest score is that the researcher can verify that randomization did produce similar groups.

Types of True Experimental Designs

A common true experimental design is the Pretest–Posttest Control Group Design (this design is sometimes referred to as the Classical Experimental Design or Randomized Before/After Control Group Design). Harris and Franklin (2003) used this design (see Exhibit 6.6a) to test whether a cognitive-behavioral group for Mexican American pregnant and parenting adolescents would improve problem solving, coping skills, and academic achievement. The experimental group participated in an 8-session task-centered, cognitive-behavioral group and received case management services, while the comparison group received only case management services.

To establish an *association,* there are two groups, one receiving the intervention and the other group receiving services as normal. You can make two different types of comparisons to demonstrate the association. If you want to assume that the Time 1 observations are more or less equal, you only need to compare O_{2a} with O_{2b}. But in this case, you are making an assumption that is not necessary; instead, it is common to compare the change between observations in Group A (ΔO_A) with the change in Group B (ΔO_B). *Time order* is demonstrated by measuring outcomes at the point of initial assignment into one of the two groups and after the intervention is provided.

To establish *internal validity* (or nonspuriousness), Harris and Franklin (2003) used a statistical package that generates random numbers to assign participants into the cognitive-behavioral group or the comparison group. Since participants were randomly assigned to the groups, any group differences should be due to chance, and there should be little to no likelihood that a third variable impacts on the findings. They did test the comparability of the two groups on a variety of demographic variables and found no differences.

The weakness of the Pretest-Posttest Control Group Design is that there might be a testing effect, that is, the pretest measurement might affect the posttest scores. It is possible that the test might sensitize respondents to the treatment and therefore make the treatment effective or more effective than it would have been without the test. The Posttest-Only Control Group Design (Randomized Control Group After-Only Design) safeguards against the reactive effects of testing and treatment. There is no pretest; rather, the researcher assumes that the pretest measures are more or less equal because random assignment is used. But because random assignment is based on probability, there is the possibility that unequal groups were formed; the lack of a pretest makes it impossible to check whether the groups are comparable initially. Walton (2001) used this design (see Exhibit 6.6b) to evaluate the effectiveness of two different methods to investigate allegations of child abuse and to develop service places. In particular, the design safeguarded against the outcomes of the assessment of parenting attitudes being influenced by a pretest measurement.

The Solomon Four Group Design enables researchers to determine whether there is a testing effect or a test-treatment interaction by combining the features of the preceding designs. There are four groups, two with a pretest and posttest and two with a posttest only. If a testing or a test-treatment interaction exists, there should be a difference in outcome scores for the two experimental groups and the two comparison groups.

Generally, this kind of interaction is problematic. When social workers apply research findings to practice, they do not want to have to recreate the research conditions. The interaction of testing and treatment

Exhibit 6.6 Examples of True Experimental Designs

a. Pretest-Posttest Control Group Design:

Effects of a cognitive-behavioral, school-based, group intervention with Mexican American pregnant and parenting adolescents (Harris & Franklin, 2003)

Subjects:	Random Assignment (R)	Pretest Measures (O_1)	Intervention (X)	Posttest Measures (O_2)
Pregnant or parenting adolescents	Group A	Social problem-solving skills Problem-focused coping School attendance Grade average	Cognitive-behavior group & case management	Social problem- solving skills Problem-focused coping School attendance Grade average
	Group B	Social problem- solving skills Problem-focused coping School attendance Grade average	Case management	Social problem-solving skills Problem-focused coping School attendance Grade average

Source: Adapted from Harris, M.B., & Franklin, C. G. (2003). Effects of a cognitive-behavioral, school-based, group intervention with Mexican American pregnant and parenting adolescents." *Social Work Research, 27,* 71–83.

b. Posttest Control Group Design:

Combining abuse and neglect interventions with intensive family preservation services: An innovative approach to protecting children (Walton, 2001)

Subjects	Random Assignment (R)	Pretest Measures (O_1)	Intervention (X)	Posttest Measures (O_2)
Families with alleged child abuse or neglect	Group A	NONE	Investigative and family preservation services	Parenting attitudes Satisfaction with services In-home status of children Number of days in-home Investigation findings Crises
	Group B	NONE	Routine services	Parenting attitudes Satisfaction with services In-home status of children Number of days in-home Investigation findings Crises

Source: Adapted from Walton, E. (2001). Combining abuse and neglect interventions with intensive family preservation services: An innovative approach to protecting children. *Research on Social Work Practice,* 11, 627–644.

might be an exception to this concern. It is common for clients to receive an initial test that we typically might refer to as an initial assessment or even a follow-up assessment. The testing effect may be beneficial if it adds to the effectiveness of the intervention.

"Difficulties" in True Experiments in Agency-Based Research

If true experimental designs are powerful tools to demonstrate causality, why are they typically the province of social work researchers and used less often by agencies to evaluate their programs? Implementing true experimental designs requires expertise, sufficient numbers of clients, and plenty of time. In addition, there are real and imagined criticisms including:

- The program cannot change during the course of the experiment or evaluation (Weiss, 1998). This poses a problem because administrators want continual improvement. Weiss recommends that if there are program changes, the timing be noted and to assess outcomes at the time when a program changes.
- Even if the program or treatment does not change, implementation depends on staff with different skill levels, such as their ability to engage clients or provide services.
- The more controlled the conditions under which the treatment or program is provided, the less generalizable it will be to other times or locations. This is certainly a concern and points to the importance of being able to describe the context and conditions under which the research is undertaken. One solution is to replicate the intervention in different sites and with different samples.
- Staff may complain about threats to professional judgment. Posavac and Carey (1997) refer to these problems as, "I know what is best for my client" (p. 184). This concern manifests itself in two ways. One concern is about random assignment. Typically, social workers choose clients for treatment based on need or when someone comes to seek help. Many social workers (including our students) object to letting chance (randomization) dictate who gets help and who has to wait; rather, they want to base their decisions on their professional judgment. The second concern is that an experiment defines the intervention that is provided to clients and how the intervention is provided. Some social workers believe this threatens their ability to make decisions about how to best meet their clients' needs.
- Staff may say, "If the experimental approach is believed to be so good, I want all my clients to get it." The reason for trying a new method is based on some belief—whether anecdotal or theoretical—that suggests that the model is indeed better. Staff want their clients to receive the best service, so they argue, "Why delay implementing what we know should work?" Yet social workers have an ethical responsibility to have evidence that the new intervention is better before it is widely implemented.
- Clients may say, "Don't experiment on me" (Posavac & Carey, 1997, p. 183). People are suspicious of experimentation, and clients, who are particularly vulnerable, may be even more suspicious of experimentation. This makes recruitment and retention more difficult. With proper human subject protections, such as the use of informed consent, these fears can be mitigated.

2 Quasi-Experimental Designs

Sometimes using an experimental design to test hypotheses about the impact of service delivery, the effectiveness of a treatment modality, or the manner in which services are provided is not feasible with the desired subjects and in the desired setting. A true experiment may be too costly or take too long to

carry out, it may not be ethical to randomly assign subjects to different conditions, or it may presume ability to manipulate an intervention that already has occurred. Researchers use quasi-experimental designs to overcome these problems.

A **quasi-experimental design** is one in which we may be able to rule out at least some alternative explanations that threaten internal validity. We will focus on two types of quasi-experimental designs, nonequivalent control group designs and time series designs.

Nonequivalent Control Group Designs

The Nonequivalent Control Group Design (see Exhibit 6.7) is exactly like the Pretest-Posttest Control Group Design except that there is no random assignment into the groups. There are two groups: One is exposed to the independent variable, whereas the other is not exposed to the independent variable. Researchers also may use this design to compare two (or more) interventions.

In this type of quasi-experimental design, a comparison group is selected as similar as possible to the experimental treatment group. This technique may employ individual matching, block matching, and aggregate matching used in true experimental designs; however, the key difference is that once the matches have been made, there is no attempt to utilize chance to randomly assign participants into the groups. The individuals or groups serving as the matched group become part of the comparison group.

Where are matches to be found? One potential source for finding matches is an agency wait-list. Persons on the wait-list are as yet not receiving services from the agency and, therefore, are a comparison group that is likely to be similar to the treatment group. Another alternative is to locate similar individuals in the community who are willing to serve as a control group. A third option is to compare client outcomes in one agency with client outcomes in another agency, assuming of course that the second agency is serving a similar population group.

Two different strategies were used in studies conducted by Wikoff, Linhorst, and Morani (2012) and Moran and Bussey (2007). Wikoff and colleagues assessed the effectiveness of a reentry program for prisoners in limiting recidivism (see Exhibit 6.7a). Unable to use random assignment, they compared released prisoners who voluntarily entered the reentry program to prisoners released at the same time who opted not to participate in the program. They acknowledged in their limitations that self-selection and therefore motivation was a potential problem (Wikoff et al., 2012). They noted that the groups differed in age, education, and gender but the groups were similar by race and other risk factors such as substance abuse, mental health, or adjustment to prison. There were unsure as to whether the several differences contributed to the outcomes.

Moran and Bussey (2007) used a similar community to establish a matched group for their study of an alcohol prevention program with urban American Indian youth (see Exhibit 6.7b). The intervention was carried out with youth living in the Denver metropolitan area and the matched group came from Colorado Springs as it was the second-largest metropolitan area with the second-largest number of American Indians in Colorado. Concerned that group differences in selection and attrition might impact the findings, they compared the two groups demographic composition and scores on various measures at the start of the study and examined whether there was differential attrition. Finding only two group differences and no differences in attrition, they were confident that the two groups were similar and their characteristics were not related to the observed group differences on the outcome measures.

Nonequivalent control or comparison group designs are particularly useful for researchers (and evaluators). Because of the pretest and posttest, both *time-order* and a *statistical association* can be demonstrated, suggesting that if not causal, there is a correlation between the treatment and outcome. Further, if the selection process appears sound, you might rule out other explanations. The key is whether you are convinced that the matched comparison group has been chosen and evaluated in such a way that you are

Exhibit 6.7 Nonequivalent Group Designs

a. Individual Matching

Recidivism Among Participants of a Reentry Program for Prisoners Released Without Supervision (Wikoff, Linhorst, & Morani, 2012)

Subjects	O_1	X	O_2 and 1-year follow-up
Group A: *Participants* Recruited inmates released from prison March 1, 2007, to February 2008 (n=108) plus 14 released June 6, 2006 to January 30, 2007.	Demographics Severity of crime Institutional risk Substance abuse Mental health	Reentry Program: Case management 6-month monetary support in form of bus passes, grocery and clothing gift cards, housing, substance abuse treatment, job training programs.	Recidivism: Conviction resulting in state-level crime to October 2009 resulting in prison or probation. Did not include fine or jail term.
Group B: *Refusal Group* Inmates released from prison March 1, 2007, to February 2008.	Demographics Severity of crime Institutional risk Substance abuse Mental health		Recidivism: Conviction resulting in state-level crime to October 2009 resulting in prison or probation. Did not include fine or jail term.

b. Community (aggregate) matching:

Results of an Alcohol Prevention Program With Urban American Indian Youth (Moran & Bussey, 2007)

Subjects	O_1	X	O_2 and 1-year follow-up
Group A: American Indian youth living in Metropolitan Denver	Demographics Impact of alcohol use Locus of control Depression Self-concept Perceived social support Decision making skills Indian identify	Culturally designed 13-week after-school program AND Six sessions six months later	Impact of alcohol use Locus of control Depression Self-concept Perceived social support Decision making skills Indian identify
Group B: American Indian youth living in Colorado Springs	Demographics Impact of alcohol use Locus of control Depression Self-concept Perceived social support Decision making skills Indian identify		Impact of alcohol use Locus of control Depression Self-concept Perceived social support Decision making skills Indian identify

Source: Walton, E. (2001). Combining abuse and neglect interventions with intensive family preservation services: An innovative approach to protecting children. *Research on Social Work Practice, 11,* 627–644.

willing to accept the comparability between the two groups despite the lack of random assignment. As you can see from the two examples, it is important to do such an assessment.

Time Series Designs

A Time Series Design is unlike the other research designs we have described up until now in that no control or comparison group is needed. A time series design typically involves only one group for which multiple observations of data have been gathered both prior to and after the intervention. Although many methodologists distinguish between Repeated Measures Panel Designs, which include several pretest and posttest observations, and time series designs, which include many (preferably 30 or more) such observations in both pretest and posttest periods, we do not make this distinction here.

A common design is the Interrupted Time-Series Design where there are three or more observations taken before and after the intervention. Like other designs, there are variations on this basic design, including time series designs with comparison or control groups and time series designs in which observations are also gathered during the course of the intervention.

One advantage of a time series design is that there is only one group, so a second group need not be created. A second advantage is that, depending on the question, both the pretest and posttest observations need not occur prospectively and, therefore, a variety of issues ranging from changes in social policy to changes in agency administrative practices to changes in agency programs can be assessed using data already collected. For example, Sandy Wexler and Rafael Engel (1999) wanted to know whether returning cash assistance policy to the states as mandated under the 1996 Federal Personal Responsibility and Work Opportunities Act would lead to what many scholars thought would be a rush to the bottom in the level of cash assistance given by the states. To examine this question, they used published data. They found that the decline in cash benefits controlling for inflation had begun many years before the passage of TANF (see Exhibit 6.8).

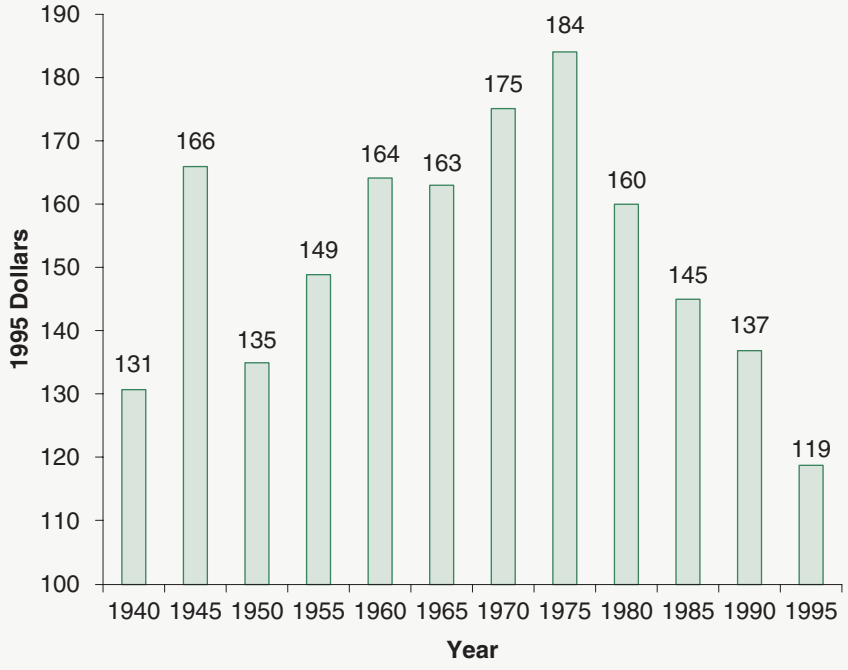
A time series design is based on the idea that, by taking repeated measures prior to an intervention or programmatic change, you have the opportunity to identify a pattern. A pattern may show a trend reflecting an ongoing increase or decline or it may stay flat. The pattern may be seasonal, with differences based on time of year such as the use of a homeless shelter that experiences declines in use during the summer and peaks in the winter, then slowly declines again. Having identified the preintervention pattern, the question is whether an intervention or program altered the nature of the pattern to what is considered to be a more favorable state. In the above example, the pattern prior to TANF reflected declining benefits and the introduction of TANF did not accelerate or alter that pattern.

What can you say about causality when using a time series design? The before-after comparison enables you to determine whether an *association* exists between the intervention and the dependent variable. You can determine whether the change in the dependent variable occurred after the intervention, so *time order* is not a problem. However, there is no control group, so spuriousness may be a problem; some other event may have occurred during the intervention that resulted in a change in posttest scores. But the multiple pretest scores do enable you to discount other explanations, as they are likely to be observed in the pretest scores.

2 Designs to Monitor Programs

Nonexperimental designs (or pre-experimental research designs) are classified as such because they provide less evidence for causality. To the extent that social work researchers are trying to demonstrate that different treatment modalities cause a change, a glaring weakness is the lack of control over threats to internal validity. This weakness often leads researchers and consumers of research to discount the utility of

Exhibit 6.8 AFDC Average Payment per Recipient in 1995 Dollars — 48 States



Source: Wexler & Engel (1999), p. 47.

these designs and the findings from studies using these designs. Yet, the simplicity of these designs makes them extraordinarily useful to evaluate agency programs.

Types of Nonexperimental Designs

The One Group Pretest-Posttest Design (Before-After One Group Design) is characterized by the absence of a comparison group; unlike the time series design, it lacks repeated pretest measures. All cases are exposed to the experimental treatment. The basis for comparison is provided by the pretest.

It is possible to demonstrate a time order because there is a pretest and a posttest. Having pre- and posttest scores means statistical analyses can be used to determine whether there is an association between the independent and dependent variables. The weakness of this design is that spuriousness is a problem; there are many different threats to the internal validity of the design.

This is a popular form of design for program evaluation both for its ease of use and for the types of questions that you might answer. It is far simpler to implement than group designs because no comparison group is needed. The design flows from a typical practice model of assessment, intervention, and evaluation of the impact of the intervention (follow-up assessment). The conformity to a practice model is more easily understood and accepted by practitioners and agency directors.

The One Group Pretest-Posttest Design provides answers to questions of interest to social service agency staff and the funders of social services. This design demonstrates whether improvement occurred, how much change occurred, and how many individuals improved. It can be used to determine how well

clients are functioning and the number of clients who have achieved some minimum standard of functioning by the program's end.

John McNamara, Kevin Tamanini, and Suzanne Pelletier-Walker (2008) used a pretest-posttest design to study the effectiveness of brief counseling with women at a domestic violence shelter (Exhibit 6.9). The social worker counselors, with a feminist treatment orientation, drew from treatment models such as cognitive-behavior, solution-focused, and family or systems counseling tailored to the individual client needs. They were interested in determining if clients who had received at least three counseling sessions would have improvements in coping ability, life functioning, symptom measures, interpersonal relationships, and carrying out social roles. McNamara and colleagues (2008) selected this design as they felt that the shelter's crisis-oriented services were not amenable to a using random assignment to a comparison group. The researchers found improvements in each of the assessed areas.

A less rigorous one group design is the After-Only Design (Posttest-Only; Cross-Sectional Group; One-Shot Only). This design is characterized by only one group without a control or comparison group, and it includes no pretest observations so that there are no benchmarks to which the posttest scores can be compared. The After-Only Design has little utility for researchers trying to illustrate causation. Because there is no pretest, both time order and association cannot be determined. The researcher does not know if the final outcomes are higher, lower, or equal to the pre-intervention level. Further, it is impossible to rule out other explanations.

The After-Only Design may be used to provide factual information for agency-based program evaluation. It is the typical design used for client satisfaction. It is also used to describe participant functioning at the end of the program, such as how many clients are no longer depressed or how many are employed after a job training program. Nonetheless, changes in depression or employment cannot be attributed solely to the program. It is also useful for researchers piloting and developing measures, developing hypotheses about relationships that then require more rigorous designs, and may provide some sense of attrition related to the treatment.

A third nonexperimental design is the Static-Group Design. It includes two groups without random assignment: One group gets the treatment and the other group does not receive the treatment, and there is no pretest or baseline. This design is frequently used when a program has already begun and baseline information cannot be obtained. The central issue of this design is finding a comparable group. If an agency wait-list is used, perhaps an argument might be made about the comparability of Group B. Or one might find nonparticipants who are eligible for the program to use as a comparison group. The problem persists that without a pretest, the comparability of the groups cannot be evaluated. Without such a test, it is a leap of faith to say that comparing posttest scores provides evidence of a time order and an association, let alone controls for internal threats to validity.

Exhibit 6.9 Pretest-Posttest Design

Pretest-Posttest Group Design:
The Impact of Short-Term Counseling at a Domestic Violence Shelter (McNamara, Tamanini, & Pelletier-Walker, 2008)

Subjects	Pretest Measures (O_1)	Intervention (X)	Posttest Measures (O_2)
Women admitted to a shelter	Life-functioning, symptoms of distress, interpersonal relationships, social roles, coping	Counseling including elements of cognitive-behavioral, existential, solution-focused, and systems	Life-functioning, symptoms of distress, interpersonal relationships, social roles, coping

Implications for Evidence-Based Research

The types of designs described throughout this chapter (see Exhibit 6.10) provide varying degrees of evidence to support the notion that a particular intervention resulted in the desired change in some outcome. There is a hierarchy among these group designs based on the three criteria for causality. True experimental designs (or randomized clinical trials) are commonly accepted as the gold standard in offering evidence about the efficacy of an intervention because they are organized to meet the criteria of association, time order, and internal validity (APA, 2006; Gambrill, 2006; Johnston et al., 2006). Ideally, there are a number of randomized experimental trials of the intervention's effectiveness relative to a particular outcome. Quasi-experimental and nonexperimental designs provide less conclusive evidence about the effectiveness of the intervention.

But, we do not mean to suggest that you need not look critically at the evidence learned from a true experiment, let alone quasi-experimental and nonexperimental designs. Throughout this chapter, we have suggested that there are specific issues you should consider as you read the results of research studies, including:

- *Randomization process.* Many authors report using random assignment of participants to the experimental and comparison groups without clarifying how the actual assignment was made. This is important information as you assess the findings' internal validity.
- *Sample size.* In Chapter 5, we briefly mentioned the concept of statistical power; the study needs to have a sample size that is sufficient to detect a statistically significant difference. With small samples, the chances of finding no treatment effect are greater than with a larger sample size; in other words, there may indeed be a treatment effect but the sample size may be too small to detect the impact of the treatment.
- *Attrition.* It is likely that some participants will drop out of the study, and there may be differential rates of attrition for the experimental and comparison groups. You should consider how attrition is handled in the analysis.

Even after you are convinced that the results are meaningful and not the outcome of a poor process of random assignment, small sample size, or attrition, you have to address the external validity of the findings. Remember, you are taking research-derived knowledge and applying that knowledge to your individual clients. Ted McNeil (2006) notes that "clinical expertise is indispensable for deciding whether external evidence applies to an individual client and, if so, how it should be integrated into treatment" (p. 151). Will the treatment's effects hold true for your clients who, for example, may differ by race, gender, social class, or sexual orientation from the people in the intervention studies? Does the study's setting and location impact on the findings? These are all considerations in determining an appropriate intervention.

Diversity, Group Design, and Evidence-Based Practice

In Chapter 5, we described how historically racial minorities and women had not been adequately represented in research studies. Under the provisions of the *NIH Revitalization Act of 1993* (PL 103-43), women

Exhibit 6.10 Summary of Group Research Designs

1. True Experimental Designs (note, there exist such as placebo designs, cross-over designs, and multiple follow-up designs)

Pretest-Posttest Control Group Design

RA	O	X	O
RB	O		O

or

Pretest-Posttest Comparison Group Design

RA	O	X_e	O
RB	O	X_t	O

Posttest-Only Control Group Design

RA	X^e	O
RA		O

or

Posttest-Only Comparison Group Design

RA	X_e	O
RA	X_t	O

Solomon Four Group Design

RA	O	X_e	O
RB	O		O
RC		X_e	O
RD			O

2. Quasi-Experimental Designs

Nonequivalent Control Group Design

A	O	X	O
B	O		O

or

Nonequivalent Comparison Group Design

A	O	X_e	O
B	O	X_t	O

Time Series (note, this design can be used with 2 or more groups)

A	O	O	O	O	X	O	O	O	O

3. Nonexperimental Designs

One-Group Pretest-Posttest Design

A	O	X	O

After-Only Posttest Design

A	X	O

Static-Group Comparison

A	X	O
		O

and minorities must be included in clinical research supported by the National Institutes of Health. We described recruitment strategies in Chapter 5. In this section, we highlight the link between adequate representation in research and evidence-based practice.

The most important consideration is the external validity of the findings from group research designs. For what population groups has the intervention been determined to be effective? Although social work research extends to many different areas, the evidence about the inclusion of people of color and women is probably best developed in the evaluations of mental health studies. Jeanne Miranda, Richard Nakamura, and Guillermo Bernal (2003) found that between 1986 and 1997 there were few minority participants in studies using true experimental designs to test the effectiveness of treatment for depression, bipolar disorder, schizophrenia, and ADHD; of nearly 10,000 participants, they could only identify 561 African Americans, 99 Latinos, 11 Asian Americans and Pacific Islanders, and no American Indians or Alaska Natives (Miranda et al., 2003; USDHHS, 2001). A more recent analysis of 379 National Institute of Mental Health-funded studies reported that women were adequately included whereas only Whites and African Americans were adequately represented (Mak, Law, Alvidrez, & Perez-Stable, 2007).

Therefore, as you review the available research and try to answer the *for whom* question, it is necessary to identify the characteristics of those who participated in the research. This is likely to be challenging because many of the studies described by Miranda et al. and Mak et al. included no information about ethnicity or lumped all those who were not White into a single category.

Representation alone is insufficient, because there needs to be a sufficient number of participants so that subgroups can be analyzed. Researchers often fail to do an analysis just for women in their sample or just for African Americans in their sample. Rather, the results are often only analyzed for the entire sample; different treatment effects for women or people of color may not be reported or observed.

Finally, the broad categories we use to depict racial or ethnic groups tend to imply that all African Americans, all Latinos, or all Asian Americans share the same cultural, social, and historical legacies. Yet, within these groups, there are differences in cultural definitions, language, history, and immigration experience. For example, Vega et al. (1998) found that Mexican immigrants have lower rates of depression than do Mexican Americans born in the United States. You can see that even within what seems like a narrowly defined ethnic group—Mexican Americans—there can be significant differences. Given that there can be so many variations, evaluating the evidence becomes even more difficult. Therefore, any intervention should at least have a theoretical base, and there is some evidence to link that theoretical base to culture (Miranda et al., 2003).

2 Ethical Issues Unique to Experimental Research

Some social science experiments involve subject deception. Primarily because of this feature, some experiments have prompted contentious debates about research ethics. Experimental evaluations of social programs also pose ethical dilemmas because they require researchers to withhold possibly beneficial treatment from some of the subjects just on the basis of chance. In this section, we give special attention to the problems of deception and the distribution of benefits in experimental research.

Deception

Deception occurs when subjects are misled about research procedures to determine how they would react to the treatment if they were not research subjects. Deception is a critical component of many social experiments, though it occurs less frequently in social work research. The problem with deception is that potential participants are not given the information they need to make an informed decision and may give consent to participate in research that they otherwise would not have agreed to with full information.

You read in Chapter 3 about Milgram's (1965) use of deception in his study of obedience to authority. Volunteers were recruited for what they were told was a study of the learning process, not a study of "obedience to authority." The experimenter told the volunteers that they were administering electric shocks to a student in the next room, when there were actually neither students nor shocks. Whether or not you believe that you could be deceived in this way, you are not likely to be invited to participate in an experiment such as Milgram's. Current federal regulations preclude deception in research that might trigger such upsetting feelings.

The overarching question is, "Is there sufficient justification to allow the use of deception?" The NASW Code of Ethics does not explicitly discuss deception in research, but it does allude to deception in discussing informed-consent:

> 5.02(g) Social workers should never design or conduct evaluation or research that does not use consent procedures, such as certain forms of naturalistic observation and archival research, unless rigorous and responsible review of the research has found it to be justified because of its prospective scientific, educational, or applied value and unless equally effective alternative procedures that do not involve waiver of consent are not feasible.

To ensure that there no harm is caused to participants, researchers use a procedure called debriefing. Debriefing involves explaining the true nature of the research to participants after the experiment to address any issues that might have arisen as a result of their participation.

Selective Distribution of Benefits

One ethical issue that is somewhat unique to experiments to assess social programs is the distribution of benefits: How much are subjects harmed by the way treatments are distributed in the experiment? For example, participation in a study of different models of case management for TANF recipients could have serious implications. The requirements of TANF impose a lifetime limit on participation so persons receiving a potentially less adequate method of case management could lose valuable time if one method helped people find work faster than the other method.

Is it ethical to give some potentially advantageous or disadvantageous treatment to people on a random basis? Random distribution of benefits is justified when the researchers do not know whether some treatment actually is beneficial—and, of course, it is the goal of the experiment to find out. Chance is as reasonable a basis for distributing the treatment as any other. Also, if insufficient resources are available to fully fund a benefit for every eligible person, distribution of the benefit on the basis of chance to equally needy persons is ethically defensible (Boruch, 1997).

12 Conclusion

True experiments play a critical role in social work research. They are the best research designs to provide evidence that an intervention is effective. Such studies are imperative for building the evidence base

of social work practice. Even when conditions preclude use of a true experimental design, many research designs can be improved by adding some experimental components. Nevertheless, the lack of random sampling limits the generalizability of the results.

Just because it is possible to test a hypothesis with an experiment does not mean it will be desirable to do so. When a social program is first being developed and its elements are in flux, it is not a good idea to begin a large research study that cannot possibly succeed unless the program design remains constant. Researchers should wait until the program design stabilizes somewhat. It also does not make sense for researchers engaged in program evaluation to test the impact of programs that cannot actually be implemented or to test programs that are unlikely to be implemented in the real world because of financial or political problems (Rossi & Freeman, 1989).

Many forms of social work research, particularly research and evaluation done in agencies, require design decisions about what is feasible. As you can see from the contents of this chapter, there are many components and factors to consider in choosing a group design. Regardless of the design used, it is important to understand the limits of the conclusions that can be made, both in terms of the internal validity of the design and the generalizability of the findings.

Key Terms

Aggregate matching
Association
Block matching
Causal effect
Comparison group
Compensatory equalization of treatment
Compensatory rivalry
Context
Control group
Diffusion of treatment
Double-blind procedures
Experimental group
External validity
History
Individual matching
Instrumentation
Internal validity
Matching
Maturation
Mechanism
Mortality
Nonexperimental designs
Nonspuriousness
Placebo effect
Posttest
Pretest
Process analysis
Quasi-experimental design
Random assignment
Reactivity
Resentful demoralization
Secular drift
Selection bias
Spurious relationship
Statistical regression
Time order
True experiments

Highlights

- Causal explanation relies on a comparison. The value of cases on the dependent variable is measured after they have been exposed to variation in an independent variable. This measurement is compared to what the value of cases on the dependent variable would have been if they had not been exposed to the variation in the independent variable.
- Three criteria are necessary to identify a causal relationship: association between the variables, proper time order, and nonspuriousness. The basis for concluding that a causal relationship exists is strengthened by identification of a causal mechanism and the context.
- Association between two variables is in itself insufficient evidence of a causal relationship.
- Threats to the internal validity of experiments include selection bias, endogenous change, external events, cross-group contamination, and treatment misidentification.

- The external validity of causal conclusions is determined by the extent to which they apply to different types of individuals and settings. When causal conclusions do not apply to all the subgroups in a study, they are not generalizable to corresponding subgroups in the population—and so they are not externally valid with respect to those subgroups. Causal conclusions can also be considered externally invalid when they occur only under the experimental conditions.
- True experimental research designs have three essential components: use of at least two groups of subjects for comparison, measurement of the change that occurs as a result of the experimental treatment, and use of random assignment.
- Random assignment of subjects to experimental and comparison groups eliminates systematic bias in group assignment. Random assignment enables researchers to presume internal validity.
- Random assignment involves placing predesignated subjects into two or more groups on the basis of chance. Matching can improve the comparability of groups when it is used to supplement randomization.
- The independent variable in an experiment is represented by a treatment or other intervention. Some subjects receive one type of treatment; others may receive a different treatment or no treatment.
- Quasi-experimental group designs control for some threats to internal validity while nonexperimental group designs tend to control for few or no threats to internal validity. It is common to find both types of designs in agency settings.
- Subject deception is common in laboratory experiments and poses unique ethical issues. Researchers must weigh the potential harm to subjects and debrief subjects who have been deceived.
- Another common ethical problem is selective distribution of benefits. Random assignment may be the fairest way of allocating treatment when treatment openings are insufficient for all eligible individuals and when the efficacy of the treatment is unknown.

Discussion Questions

1. A program has recently been funded to provide casework intensive services to the homeless. The mission of the program is to provide skills that lead to self-sufficiency and employment. Develop a research study using:

   Experimental design

   Quasi-experimental design

   Pre-experimental design

   Be specific in describing the procedures you would have to do to implement your design. This may mean specifying how you will assign clients to groups (if you have more than one group) or where you would find clients for your control/comparison groups (if you have such groups).
2. Identify the benefits and weaknesses of each of the specific designs you chose for Exercise 1.
3. Search for a research study using a true experimental design to examine the effects of hospice care. Diagram the experiment using the exhibits in this chapter as a model. How generalizable do you think the study's results are to the population from which cases were selected? To other populations? To specific subgroups in the study? To other settings? How thoroughly do the researchers discuss these issues?

Critiquing Research

1. Go to the book's study site, http://www.sagepub.com/fswr2e and choose two research articles that include some attention to causality (as indicated by a check in that column of the article matrix). For each article describe the following:

   a. What type of design was used? How does the author describe the design? Was it suited to the research question posed and the specific hypotheses tested, if any? Why do you suppose the author chose the particular design?
   b. Did the design eliminate threats to internal validity? How did the design do this? Are you satisfied with the internal validity conclusions stated by the author? Why or why not?
   c. What is the setting for the study? Does the setting limit the generalizability of the results to other similar settings or to the broader population? Is reactivity a problem? Are there other threats to external validity?
2. Find a research study reported in a newspaper or magazine. Describe what you believe are the most likely sources of internal invalidity.

Making Research Ethical

1. What specific rules do you think should guide researchers' decisions about selective benefits?
2. Under what conditions do you think that randomized assignment of subjects to a specific treatment is ethical in social work research? Was it ethical for Walton et al. (1993) to randomly assign individuals accused of domestic violence to an arrest or nonarrest treatment? What about randomly assigning some welfare recipients to receive higher payments than others?
3. Critique the ethics of one of the experiments presented in this chapter. What specific rules do you think should guide researchers' decisions about subject deception and the selective distribution of benefits?

Developing a Research Proposal

If you are planning to use a group design:

1. What specific design will you use? How long will the study last? At what time points will data be collected? How will the data be collected?
2. If you are using a design with more than one group, describe how participants are assigned to each group.
3. Discuss the extent to which each source of internal validity is a problem in the study.
4. How generalizable would you expect the study's findings to be? What can be done to improve generalizability?
5. Develop appropriate procedures for the protection of human subjects in your study. Include in these procedures a consent form.

Web Exercises

1. Try out the process of randomization. Go to the website www.randomizer.org. Type numbers into the randomizer for an experiment with 2 groups and 20 individuals per group. Repeat the process for an experiment with 4 groups and 10 individuals per group. Plot the numbers corresponding to each individual in each group. Does the distribution of numbers within each group truly seem to be random?

2. Participate in a social psychology experiment on the Web. Go to www.socialpsychology.org/expts.htm. Pick an experiment in which to participate and follow the instructions. After you finish, write up a description of the experiment and evaluate it using the criteria discussed in the chapter.

STUDENT STUDY SITE

Visit **www.sagepub.com/fswr2e** to access additional study tools including eFlashcards, web quizzes, web resources, interactive exercises, data sets, links to SAGE journal articles and more.