PART I

Critical Appraisal of Research to Support Scholarship

CHAPTER 1

Quantitative Research

■ Robin Lennon-Dearing and Susan L. Neely-Barnes ■

Critical thinking involves the careful examination and evaluation of beliefs and actions. . . . The danger of thinking uncritically is that time and resources may be wasted – or worse, that clients won't get the help they need.

— Leonard Gibbs & Eileen Gambrill, 1996

■ Objectives:

- Identify steps in the quantitative research process.
- Identify preexperimental, quasi-experimental, and experimental research studies when examining published research.
- Assess internal and external validity of various research designs.
- Recognize and understand the methodological issues in quantitative research designs.

■ Critical Appraisal

The goal of this chapter is to help readers understand the process of quantitative research so they can critically identify the usefulness of different studies for their own research or clinical practice. Appraising information critically and in a systematic way is important to practitioners' ability to base their clinical decisions on the research evidence. Healthcare providers must understand the basic process of quantitative research to distinguish the strengths and weaknesses of a study they may be evaluating.

■ Quantitative Research

Quantitative research involves a systematic process, the scientific method, to build knowledge. Quantitative research methods involve collecting numerical data to explain, predict, and/or control phenomena of interest. Data analysis is mainly statistical; it answers questions of what, and under what condition(s),

specific independent variables predict or explain dependent variables through the use of numerical data suitable for statistical analysis (Solomon & Draine, 2010). Depending on the problem or issue under inquiry and after researchers have identified sufficient knowledge from a literature review, they begin with a research question or hypothesis (Keele, 2011). Whereas quantitative research questions look at the relationships among variables, quantitative hypotheses are predictions the researcher makes about the expected relationship among variables. The research design becomes the blueprint for the study—that is, how the study sample is selected and how the data are collected and analyzed (Keele, 2011). An overview of the basic steps in the quantitative research process is shown in **Table 1-1**.

When a problem of interest has been identified, the research process is applied to discover what is known about a topic and where knowledge gaps exist (Schmidt & Brown, 2012). The researcher then finds existing knowledge on a subject from a review of relevant literature. From what is learned in relation to the research problem from the literature review, a focused research question should follow (Yegidis & Weinbach, 2009). **Table 1-2** shows how the problem of interest has been narrowed to an answerable question and then to a hypothesis statement. A research hypothesis is stated as an answer to a research question (Yegidis & Weinbach, 2009).

The research hypothesis commonly states the type of relationship, as described in **Table 1-3**, between variables that it is presumed they have. Objective measurable data are then collected to confirm or refute a hypothesis (Schmidt & Brown, 2012).

In quantitative research studies, variables are numerical (Brown, 2012). Biophysical variables such as height, weight, blood pressure, and pulse may be measured directly. Conceptual variables have attributes or characteristics that differ in quantity or quality and describe people or things (Babbie, 2010), and they must be operationalized—that is, defined in terms that give precise indicators to be observed, and

Table 1-1 Steps in Quantitative Research

1. Problem identification
2. Research question formulation
3. Literature review
4. Construction of hypothesis
5. Research design and planning
6. Data collection
7. Sorting and analysis of data
8. Specification of research findings
9. Interpretation of research findings
10. Dissemination of research findings
11. Use of findings by practitioner

Source: Yegidis & Weinbach, 2009.

Table 1-2 Study Example of a Research Question and a Research Hypothesis

Study	*Research Question*	*Research Hypothesis*
Demark-Wahnefried, W., Hars, V., Conaway, M. R., McElveen, G., & Winer, E. P. (1997). Reduced rates of metabolism and decreased physical activity in breast cancer patients receiving adjuvant chemotherapy. *American Journal of Clinical Nutrition,* 65, 1495–1501.	Do changes in energy intake, physical activity, resting metabolic rate (RMR), diet-induced thermogenesis, or any combination of these variables occur that may contribute to weight gain in women receiving adjuvant chemotherapy for breast cancer?	A reduction in RMR would be observed during the period in which women received adjuvant chemotherapy.
Erblich, J., Boyarsky, Y., Spring, B., Niaura, R., & Bovbjerg, D. H. (2003). A family history of smoking predicts heightened levels of stress-induced cigarette craving. *Addiction,* 98, 657–664.	What differences between smokers with and without histories of smoking in first-degree relatives might explain the risk for persistent smoking and relapse?	Smokers with two or more first-degree relatives who smoked would exhibit stronger craving reactions following stressful stimuli than smokers without such family histories.

specify the level of those indicators (Rubin & Babbie, 2011). Tools used to measure conceptual variables are called instruments.

As shown in **Table 1-4**, the independent variable is what the researcher introduces and controls to measure its effect on the dependent variable (Yegidis & Weinbach, 2009). The dependent variable is the focus of the intervention and is what is measured. Confounding variables are factors that interfere with the relationship between the independent and dependent variable (Schmidt & Brown, 2012).

Research hypotheses suggest and test for relationships between variables. Relationships between variables can be positive, negative (inverse), or curvilinear. For example, in a study looking at the role of social networks and support as they relate to symptoms of depression in women who have recently given birth, Surkan, Peterson, Hughes, and Gottlieb (2006) chose the Medical Outcomes Study Social Support Survey and a social network item as the independent variable, and the Center for Epidemiologic Studies of Depression Scale as the dependent variable. Using the

Table 1-3 Relationships Between Variables Expressed in Hypotheses

Association	Certain value categories of X are found with certain value categories of Y.
Correlation	Higher values of X are found with higher values of Y and vice versa, or higher values of X are found with lower values of Y and vice versa.
Causation	Values or value categories of X cause values or value categories of Y.

Source: Yegidis, B. L. & Weinbach, R. W., 2009. *Research methods for social workers* (6th ed.). Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.

Table 1-4 Types of Variables

Independent Variable	This is manipulated by the researcher to influence the dependent variable; may also be called predictor variable.
Dependent Variable	This is the variable of primary interest to the researcher; may also be called outcome variable.
Confounding Variable	An extraneous third variable that influences the relationship between the independent and dependent variables.

Source: Yegidis & Weinbach, 2009.

appropriate statistical analysis, the researchers found that both social networks and social support were independently and inversely correlated to symptoms of depression. Women who reported more social support from friends and family showed fewer depressive symptoms and reported lower scores on the measure for depression.

The strength and direction of a relationship, the *effect size*, between two variables can be statistically tested and reported using a correlation coefficient, such as Pearson's *r*. The direction of the relationship is positive (+1.0 is a perfect positive relationship) or negative (–1.0 is a perfect negative relationship). The closer the value gets to +1 or –1, the stronger the relationship; a value close or equal to 0 indicates no relationship (Brown, 2012). High correlation only implies a pattern in the relationship between variables; it does not equal causation (Brown, 2012).

■ Sampling

To answer the research question and test the research hypothesis, a researcher must define the population of interest. Studying an entire population of interest is usually prohibitive in terms of time, money, and resources, so a subset of a given population must be selected; this is called sampling (Yegidis & Weinbach, 2009). The method used for choosing a sample affects its representativeness of the population and thus the generalizability of results. There are two types of sampling: *probability* sampling and *nonprobability* sampling. Probability sampling means that "all members of that population have an equal chance of being selected in the sample" (Rubin & Babbie, 2011, p. 360). The four probability sampling methods (see **Table 1-5**) are: simple random sampling, stratified sampling, cluster sampling, and systematic sampling (Schmidt & Brown, 2012).

Nonprobability sampling (see **Table 1-6**) uses methods such as convenience sampling, quota sampling, purposive sampling, and snowball sampling (Schmidt & Brown, 2012). For some research studies, probability sampling is not possible or not feasible because of costs. In these situations, the researcher must rely on nonprobability methods. Research studies that use nonprobability methods can have scientific merit but will have limited generalizability to the larger population.

Table 1-5 Probability Sampling Methods

Method	*Definition*	*Benefits and Limitations*
Simple random sampling	Each subject has the same chance to be selected. Strategy used upholds randomization.	High probability that the sample will represent the population as long as sample size is sufficient.
Stratified random sampling	Strata must be mutually exclusive so a subject can be assigned to only one stratum. Random sampling used to select subject from each stratum.	High probability that the sample will represent the population if number of subjects in each stratum is sufficient.
Cluster sampling	Simple random sampling used first to select clusters and then select subjects within each cluster.	Greater potential for the sample to not represent the population depending on how the initial clusters are selected.
Systematic random sampling	Begin with random sampling and count the Nth subject on the list.	If bias occurs, this type of sampling is not as representative as the other three methods.

Sources: Adapted from (1) Haber, J. (2006). Sampling. In G. LoBiondo-Wood & J. Haber (Eds.), *Nursing research: Methods and critical appraisal of evidence-based practice* (pp. 121–143). Sudbury, MA: Jones and Bartlett; and (2) Wood, M., & Ross-Kerr, J. (2006). *Basic steps in planning nursing research: From question to proposal* (6th ed.). Sudbury, MA: Jones and Bartlett.

Table 1-6 Nonprobability Sampling Methods

Method	*Definition*	*Benefits and Limitations*
Convenience sampling	Inclusion criteria identified prior to selection of subjects. All subjects are invited to participate.	Because the sample is selected for ease of data collection, it may not be representative of the target population.
Quota sampling	Strata must be mutually exclusive so a subject can be assigned to only one stratum. Convenience sampling used to select subject from each stratum.	Because the sample within each stratum is selected using convenience sampling, it may not represent the population.
Purposive sampling	Researcher has sufficient knowledge of topic to select sample of experts. Researcher should identify criteria to include in selection of subjects.	Because the sample is selected by researcher, cannot generalize to population; generalizing the results is not an expected outcome.
Snowball sampling	Researcher selects initial subjects for study. Data saturation is reached.	Cannot generalize to population; generalizing the results is not an expected outcome.

Sources: Adapted from (1) Haber, J. (2006). Sampling. In G. LoBiondo-Wood & J. Haber (Eds.), *Nursing research: Methods and critical appraisal of evidence-based practice* (pp. 121–143). Sudbury, MA: Jones and Bartlett; and (2) Wood, M., & Ross-Kerr, J. (2006). *Basic steps in planning nursing research: From question to proposal* (6th ed.). Sudbury, MA: Jones and Bartlett.

■ Data Collection

Quantitative data collection methods rely on structured data collection instruments that produce results that are easy to summarize, compare, and generalize. Four levels of measurement are used to quantify data, depending on what is being measured. Nominal measures differentiate between categories but do not place variables in any order or ranking. Ordinal measures rank categories in order but do not specify the distance between the categories. Interval measures use continuous data in which values are rank-ordered, and the distance between categories is equal. Ratio scales, the highest level of measurement, measure equal interval data and employ a fixed-point zero (Schmidt & Brown, 2012).

Common data collection methods of quantitative research include questionnaires, rating scales, and physiologic measures such as blood tests and vital signs (Keele, 2011). In this chapter, we provide a basic overview of issues of validity (see **Table 1-7**) and reliability (see **Table 1-8**)of measure. Readers are encouraged to consult other texts for in-depth reviews of measurement construction and measurement theory.

Reliability

Reliability measures the consistency and stability of responses over time in a standardized measurement instrument. Reliability does not ensure that measures are accurately measuring what researchers think they measure (Babbie, 2010). *Internal consistency reliability* is a measure of how closely items in a questionnaire measuring the same construct are related. Cronbach's alpha addresses overall average reliability, and items are considered to represent a similar construct when alpha is approximately 0.80.

Table 1-7 Measurement Validity

Construct	It is *convergent* when results correspond to the results of methods measuring the same concept. It has *discriminant* validity when results do not highly correspond to other constructs as they do with measures of the same construct.
Content	Experts judge whether the measure covers the range of meanings within the concept.
Criterion-related or concurrent	Compares with an external measure of the same variable.
Face	Appears to measure what the researcher intended.
Factorial	How many different constructs are measured and whether these are what the researcher intends to measure.

Source: Rubin & Babbie, 2011.

Table 1-8 Reliability

Interrater reliability	The degree of agreement or consistency between raters.
Test-retest reliability	A measure that provides consistency in measurement over time.
Internal consistency reliability	This assesses the correlation of scores on each item with the scores on the rest of the items. Cronbach's alpha should have a value of 0.80 or greater to be considered reliable.

■ Research Design

The value of evidence from a study depends on the design used. In quantitative research, a clearly defined step-by-step process is followed based on the research design chosen (Schmidt & Brown, 2012). The following pages review research designs (see **Table 1-9**) that are used as tools to answer research questions and test research hypotheses.

■ Group Design

Group design is a commonly used technique in quantitative research and relatively well-known among students of research. When asked to design a research study, most students of quantitative methods will incorporate a group design. Group design is defined by Grinnell and Unrau (2011, p. 565) as "research design conducted with two or more groups of cases, or research participants, for the purpose of answering research questions or testing hypotheses." The method encompasses preexperimental, quasi-experimental, and experimental techniques. The most rigorous of group designs have an explanatory purpose to prove cause-effect relationships, whereas the least rigorous of these designs are used to generate or explore a theory.

There are many variations of group design. The more commonly used designs will be covered. Readers are encouraged to consult other texts for a more in-depth review.

Table 1-9 Research Types

Exploratory research	Preexperimental	Research is conducted to explore a topic about which little is known.
Descriptive research	Quasi-experimental	Descriptive research involves collecting data to test hypotheses or answer questions concerning the current status of the subjects of the study. Describes the variables. Lacks the element of random assignment.
Explanatory research	Experimental	Participants are assigned to groups based on some selected criterion often called an independent variable. At least one variable is manipulated so as to measure its effect on one or more dependent variables.

Internal Validity

From the evidence-based practice perspective, rigorous group designs are more valued than less rigorous designs. This is because rigorous designs minimize threats to internal validity. Readers should remember that internal validity is concerned with the possibility that a change in the dependent variable (outcome) is the result of some other cause than the independent variable that is the target of the experiment. It is beyond the scope of this chapter to include an in-depth review of all threats to internal validity. Briefly, one should remember that respondents improve for many reasons other than the intervention or technique that is the target of the research experiment. It is possible that research subjects improve because they age (maturation); because they can better fill out the measure of the dependent variable (testing); or because they are exposed to an external event that caused the improvement (history). It is also possible that research subjects would have improved regardless of the experimental intervention (regression to the mean), or for other reasons not mentioned here.

Whereas internal validity refers to the confidence with which the study results can conclude that a treatment or intervention (independent variable) causes change in the dependent variable (see **Table 1-10**), external validity has to do with the

Table 1-10 Internal Validity

Threats to Internal Validity	*Maximizing Internal Validity*
Internal validity is the degree to which we can confidently conclude that the treatment caused the outcomes observed.	
History–*Events occurring between repeated measurements.*	Use a control group from the same population as the experimental group.
Maturation–*Changes in participants that occur over time.*	Use a control group and keep the study of short duration.
Testing–*Change resulting from being measured; practice effect.*	Use a research design that does not include a pretest or unobtrusive data collection.
Instrumentation–*Changes in outcome because of equipment or human factors.*	Use standardized instruments, administration, or data collection procedures.
Statistical regression–*The natural tendency of very high or low scores to regress toward the mean during retest.*	Avoid using extreme scores.
Mortality–*Participants dropping out.*	Use random assignment with large groups and follow up with a portion of those who leave the study.
Selection of subjects–*Choosing participants in such a way that groups are not equal before the experiment.*	Use random selection and random assignment of subjects. If random selection and assignment are not possible, use certain other statistical techniques.

Source: Rubin & Babbie, 2011.

generalizability of the research findings. Rubin and Babbie (2011) described external validity as "the extent to which we can generalize findings of a study to settings and populations beyond the study conditions" (p. 247). They also noted that "a study must be generalizable to some real-world settings." Characteristics of good quantitative research are presenting the research design and methods in enough detail that other researchers could replicate the study and obtain their own results (Durbin, 2004). Obtaining the same results through repeated experimentation by different researchers increases the value and worth of the findings (Durbin, 2004).

Preexperimental Design

The purpose of preexperimental designs is to explore new topics of research. Preexperimental designs rank low in the evidence-based practice hierarchy (Rubin & Babbie, 2011). Yet, the designs have an important role in testing new intervention approaches, evaluating programs, and generating theories. Examples of research questions that could be addressed using a preexperimental design include: (1) Are patients leaving the hospital satisfied with discharge planning services? (2) Are patients in a health education program doing better than they were before they started?

One-Shot Case Study

The one-shot case study is the most basic of group designs, so it is a good starting point. However, it is a weak design. Campbell and Stanley (1963) noted that these studies have a total absence of control and almost no scientific value. One-shot case studies are usually diagrammed as follows, with X standing for a stimulus such as an intervention, and O standing for an observation.

X O

Despite the weakness of this study design, one-shot case studies are used quite frequently. In higher education, student evaluations of teaching are an example of this design. Many hospitals and social service agencies use this design to ask patients or participants about their knowledge or skills gained from a service. The problem with this design is that there are no points of comparison. We do not know the respondents' level of knowledge or skills prior to receiving the service, nor do we know how their current level of knowledge or skills compared with those individuals who did not receive services. Many other options are available to provide a more rigorous design.

One-Group Pretest-Posttest

The one-group pretest-posttest design assesses the dependent variable before and after the stimulus or intervention is introduced. It is usually diagrammed as follows (Campbell & Stanley, 1963):

O_1 X O_2

This design has the advantage of establishing both time ordering and correlation. A researcher can use this design to demonstrate that the study group improved if

scores are better at Observation 2 than they were at Observation 1. For reasons related to internal validity, this design cannot establish causality. For example, imagine that you are evaluating a diabetes education program for adolescents aged 12–15 years. You hypothesize that the program will improve healthy eating habits and reduce blood glucose levels. The program lasts for 1 year. You give a pretest at the beginning of the year and a posttest at the end of the year. You are able to establish that the adolescents' eating habits and blood glucose levels have improved. Did your program cause the change? There are several alternative explanations: (1) It could be that the adolescents' eating habits and management of their blood sugar improved because the adolescents matured and were 1 year older at the time of the posttest. (2) It could be that something extraneous occurred during that year that caused the change. For example, a popular show geared toward teens portrayed a young adult with diabetes. (3) It could be that the adolescents were referred when they were at their worst period of management, and they would have improved anyway. Without the presence of a control group, it is not possible to rule out these alternative explanations.

Quasi-Experimental Design

There are many situations in which it is not possible for researchers to use experimental designs. It may be unethical to deny treatment to a control group. Agency or hospital administration may not allow program participants to be randomly assigned. In these situations, quasi-experimental designs can be used. Quasi-experimental designs usually involve assignment to two groups without randomization or the use of a comparison group in place of a control group. Although less rigorous than an experimental design, quasi-experimental designs are an improvement over preexperimental designs. Three common quasi-experimental approaches will be reviewed here. Readers interested in a more in-depth discussion of the approach should consult other texts (Cook & Campbell, 1979).

Nonequivalent Comparison Groups

Suppose that one high school in town has adopted a novel sex education curriculum. You as a researcher would like to evaluate this curriculum as compared with the usual one, but the principal will not allow any students to be assigned to a control group. However, a high school across town has similar demographics to the one with the novel curriculum. The principal of this high school agrees to participate in your study and have students fill out the same pretest-posttest as the high school with the novel curriculum. In this example, you have a quasi-experimental design with nonequivalent comparison groups. You are not able to randomly assign the students to their conditions, but you hope that the two groups are similar enough to be comparable. This design is denoted:

$$O_1 \quad X \quad O_2$$
$$O_1 \qquad O_2$$

This use of the comparison group in this design addresses the concerns that students might have changed because of aging or an external event. Yet, some problems still remain in this design. The two groups were not randomly assigned. If their outcomes are different, we cannot rule out the possibility that demographic differences between the groups led to the change. Additionally, the comparison group is not a true control group. If the two groups have the same outcomes, we will be able to say that neither is superior, but we cannot answer the question of whether either approach is better than no education.

Time-Series Design

As mentioned, one concern in experimental research is that the intervention group may have changed regardless of the intervention. One of the ways of examining whether this is true is to administer multiple pretests before starting the intervention. By using multiple pretests, the researcher can detect whether there was a trend. In other words, was the group already engaged in a change process before the intervention started?

A more rigorous extension of the multiple pretest design is a time-series design. The time-series design allows the research to examine the question of whether there was a trend in the data both before the intervention and after. Opinions differ as to how many pretests and posttests are needed in a time-series design. In the example that follows, the dependent variable is measured four times before the intervention and four times after:

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$$

To further increase the rigor, researchers can use a multiple time-series design. The multiple time-series design adds a nonequivalent comparison group. The nonequivalent comparison group gets the same number of observations of the dependent variable in the same time frame but does not receive the intervention. The multiple time-series design addresses the concern that an external event occurring simultaneous to the intervention could have influenced the dependent variable. It is usually denoted:

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$$
$$O_1 \quad O_2 \quad O_3 \quad O_4 \qquad O_5 \quad O_6 \quad O_7 \quad O_8$$

■ Case Control Studies

Many questions do not lend themselves to experimental designs. Suppose we want to understand what leads a person to become a perpetrator of child abuse, what contributes to becoming a high school dropout, or what health habits contribute to high blood pressure. Designing a controlled experiment to answer one of these questions may be difficult or even impossible. Though not as rigorous as an experimental design,

a case control study is a good alternative. A case control study collects retrospective data from people who are and are not in the outcome condition and uses multivariate statistical analysis to compare the two groups and identify variables that may have contributed to the outcome condition. It is a more convenient and inexpensive way to collect outcome data than an experimental design. A downside of this design is that it relies on retrospective data. Some participants may have difficulty recalling events and circumstances of their early life, and many may not recall accurately.

Experimental Design

Experimental designs seek to answer explanatory research questions. In explanatory research, the investigator seeks to test hypotheses and explain how an independent variable influences a dependent variable. In an ideal experiment, it would be possible to say with certainty that an independent variable caused a dependent variable. It is unusual for a researcher in nursing or any medical or social science field to have sufficient control over the design of an experiment to produce the ideal (Grinnell, Unrau, & Williams, 2011). Yet, there are three criteria that can produce a high degree of certainty that an explanatory relationship exists (Rubin & Babbie, 2011):

1. The independent variable (cause) should come before the dependent variable (effect) chronologically.
2. The independent and dependent variables should be empirically related to each other.
3. The relationship between the independent and dependent variables cannot be explained as the result of the influence of a third variable.

Two key techniques in experimental design separate it from preexperimental or quasi-experimental design. The first is the use of a *control group*. A control group is a set of research respondents who resemble the experimental group in every way except that they do not receive the target intervention of the research study (Rubin & Babbie, 2011). The second technique is *randomization*. Randomization is the assignment of respondents to either the experimental or control group at random. Techniques for randomization include flipping a coin, using a random numbers table, and assigning by an even or odd identification number (Rubin & Babbie, 2011). Without randomization, there is a chance that participants assigned to either an experimental or control group could be inherently different from each other. In other words, there is a risk of *selection bias*. The term *randomized controlled trial* used frequently in evidence-based practice refers to experimental group designs with both randomization and a control group. Three of the designs most commonly discussed in the research literature are reviewed here (see **Table 1-11**).

Pretest-Posttest Control Group Design

The first type of experimental design, sometimes known as the classic experimental design, is denoted as follows, with R signifying randomization to group:

R	O_1	X	O_2
R	O_1		O_2

The classical experimental design minimizes many threats to internal validity, including maturation, history, and selection bias. This design does not account for the problem of testing effects. It is possible that participants in both the experimental and control groups will improve simply because they are retested on the same measure and have improved in completing the measure. To address the problem of testing, a different design will be described next.

Solomon Four-Group Design

If researchers would like to know about pretest-posttest change but are concerned about the problem of testing effects, they can use the Solomon four-group design. This is a highly regarded research design that involves dividing respondents into four groups: two are experimental, and two are control. One of the experimental groups and one of the control groups are pretested but not the other. It is denoted:

R	O_1	X	O_2
R	O_1		O_2
R		X	O_2
R			O_2

Alternative Treatment Design or Dismantling Study

Researchers often seek to compare alternative treatment approaches. For example, researchers may want to compare two drugs, two patient education programs, or two case management strategies. One method of comparing is to randomly assign participants to one of two groups: one receiving intervention A (X_A) and one receiving intervention B (X_B). Such a design could answer which of the two treatment alternatives is superior. However, what if the researcher is concerned that both treatments have no effect? To answer this question, a control group must be included in the study design. Then, the study would consist of three groups: one receiving intervention A, one receiving intervention B, and a final receiving no intervention. This would be denoted:

R	O_1	X_A	O_2
R	O_1	X_B	O_2
R	O_1		O_2

A final design called a dismantling study can be used to explore which components of the intervention are needed to achieve the desired effect. In the first group, participants are randomly assigned to receive both intervention components A and B. In the second, participants receive only intervention A. In the third, participants

receive only intervention B. The final group is a control group receiving no intervention. If either of the groups in the second or third rows shows as much improvement as the first group, the component in the second or third row would be all that is needed (Rubin & Babbie, 2011). This approach is denoted:

R	O_1	X_{AB}	O_2
R	O_1	X_A	O_2
R	O_1	X_B	O_2
R	O_1		O_2

An example of a dismantling study can be found in an article by Kroeze, Oenema, Dagnelie, and Brug (2008). This study examined a computed-tailored intervention aimed at reducing dietary fat intake among adults. The four conditions in the dismantling study were: (1) feedback on dietary fat intake, (2) feedback relative to one's peers, (3) the first two types of feedback plus practical suggestions on how to change fat intake, and (4) general information. Kroeze and colleagues found that the third condition, personal and peer feedback with practical suggestions, was effective in reducing fat intake among the high-risk populations. The first two conditions were only effective in changing intention to reduce fat intake.

Reactivity and Placebo Effects

All the experimental designs described earlier involve the use of a control group. The use of a control group introduces rigor in a study design to address many threats to internal validity. However, it also introduces problems of reactivity of study participants. It is possible that experimental group participants will improve simply because they are receiving additional attention that accompanies treatment. Another possibility is that control group participants will become frustrated with the study because they are not receiving treatment and drop out. On the other hand, control group participants may engage in compensatory rivalry, trying to find treatments elsewhere that mirror the one that the experimental group is receiving. All these possibilities threaten the validity of the study.

One option to address reactivity is to use a placebo. Use of a placebo has become standard practice in drug studies, but it can also be used in other types of intervention studies. Researchers who examine psychosocial or health education interventions may be concerned that the additional time and attention given to the experimental group over the control group will influence the outcome regardless of whether the intervention is effective. Thus, some researchers will introduce an alternative program for the control group that is not believed to impact the dependent variables of interest. For example, Duru, Sarkisian, Leng, and Mangione (2010) completed a randomized controlled trial of a faith-based physical activity intervention for older African American women. Because the researchers were concerned about placebo effects, the control group received group lectures about topics important to seniors, such as financial

Table 1-11 Study Examples of Research Designs

Study	*Research Design & Sampling*	*Instruments*	*Intervention*	*Findings*
Wyatt, T. H., & Hauenstein, E. J. (2008). Pilot testing Okay With Asthma: An online asthma intervention for school-age children. *Journal of School Nursing, 24*(3), 145–150.	**One-group pretest-posttest** quasi-experimental design; convenience sample	The Asthma Information Quiz; The Child Attitude Toward Illness Scale Given at baseline and 1 week and 2 weeks after the intervention	Okay With Asthma program	Significant improvements in asthma knowledge scores at the 1- and 2-week evaluations and significant improvements in attitude scores 2 weeks after the program.
Bjorkman, T., & Hansson, L. (2007). Case management for individuals with a severe mental illness: A 6-year follow-up study. *International Journal of Social Psychiatry, 53*(1), 12–22.	**Time series design**; clients were interviewed at admission, at an 18-month follow-up, and at a 6-year follow-up No control group	Lancashire Quality of Life Profile (initial); Manchester Short Assessment of Quality of Life (follow-up); Interview Schedule for Social Interaction; Strauss Carpenter scale; Camberwell Assessment of Needs interview; Hopkins Symptom Check List-90; Satisfaction with case management services	Case management services	Decrease in the use of psychiatric services and sustained improvements in social functioning.
Swenson-Britt, E., Carrougher, G., Martin, B., & Brackley, M. (2000). Project Hope: Changing care delivery for the substance abuse patient. *Clinical Nurse Specialist, 14*(2), 9–100.	**Solomon-Four design**; 80 nurses (20 per group) from four units were randomized into experimental and control groups	Questionnaires contained four sections: Section 1 assessed their personal knowledge; Section 2 assessed attitudes of nurses about the use of alcohol and alcoholism; Section 3 tested pathophysiology and medical-surgical nursing standards of care; and Section 4 asked for demographic information and previous educational or personal experience with alcoholism.	Four-session educational intervention	After the intervention, nurses' knowledge increased, but their attitudes did not change significantly.

planning. These group lectures were useful to the participants but were not expected to impact the outcome variables, such as body mass index and blood pressure.

■ Systematic Reviews and Meta-Analyses

From an evidence-based practice perspective, systematic reviews and meta-analyses hold the spot at the top of the hierarchy of research evidence. The purpose of systematic reviews and meta-analyses is to create an unbiased synthesis of the literature on a particular research question. The terms *systematic review* and *meta-analysis* are not synonymous, but the two techniques are highly compatible and can be used together to summarize a large body of research and generate new insights (Littell, Corcoran, & Pillai, 2008).

For example, Shah and Shah (2010) were interested in whether domestic violence during pregnancy has an adverse impact on the fetus. A literature review turned up a large number of studies. Some of the studies found that domestic violence increases risk, and others found no impact. How does one make sense of this variation in the literature? Shah and Shah used the systematic review process to search for literature and evaluate it. They used meta-analysis techniques to combine the results of multiple studies. Their conclusion was that domestic violence is associated with increased risk of low birth weight and preterm birth.

Systematic Review

A systematic review is a process of comprehensively locating and synthesizing the research on a particular question using organized, transparent, and replicable procedures (Littell, et al., 2008). The first step in the systematic review process is to develop a protocol. The first element of a protocol is a clearly formulated and answerable research question and a set of hypotheses. As part of the research question, there should be explicit inclusion and exclusion criteria to determine which studies are to be included in the review. These inclusion/exclusion criteria will specify problems or conditions, populations, interventions, settings, comparisons, outcomes, and study designs that are or are not to be included in the review. The protocol will specify the techniques to locate and screen studies. These techniques include search terms, databases and search engines to be used, and strategies to locate unpublished studies. When a systematic review is being prepared for inclusion in the Cochrane or Campbell Library, the protocol is submitted to and approved by peer review before the systematic review process begins. The final version of the approved protocol is posted online (Higgins & Green, 2011).

After the protocol has been formulated, the researchers locate and screen studies. Ideally, the researchers should keep a record of every abstract screened and the method by which it was retrieved. Database searches are usually the first step in a systematic review. Many systematic reviews will augment the database search with a hand search of 10–15 journals that frequently publish on the topic of review. Strong reviews will make every effort to locate unpublished studies. Methods for

finding unpublished studies include reviewing proceedings of relevant conferences and searching the websites of government and nonprofit organizations that have an interest in the study topic. After the initial screening, two reviewers will read the study and determine whether it meets eligibility criteria for inclusion in the review. If the two reviewers disagree, a third usually breaks the tie.

After studies are located and screened, included studies are rated for study quality, and data are extracted from the study. Data extraction involves recording the sample size and characteristics, the type of interventions used (if the focus of the research question is intervention), and the outcome variables and measures chosen. Study quality ratings are undertaken to assess whether there is any bias in the reporting of study outcomes. The Cochrane Handbook (Higgins & Green, 2011) recommends that reviewers assess the following types of bias: (1) selection bias—whether there were systematic differences in the composition of groups; (2) performance bias—whether there were systematic differences in care between the groups other than the intervention; (3) attrition bias—whether one group withdrew or dropped out at a higher rate than the other; (4) detection bias—whether there were systematic differences in outcome assessment because of unblinded assessment; and (5) reporting bias—whether there was a tendency to report only significant findings.

Meta-Analysis

Meta-analysis has been defined as "a set of statistical techniques for combining quantitative results from multiple studies to produce a summary of empirical knowledge on a given topic" (Littell et al., 2008, pp. 1–2). Meta-analysis is used after data have been extracted in the systematic review process. A meta-analysis produces an effect size, a measure of strength and direction of a relationship. Several different metrics can be used to estimate the effect size in a meta-analysis. When dependent variables are continuous, it is common to use standardized mean differences, also known as Cohen's *d*. When dependent variables are dichotomous, odds ratios or risk ratios are frequently the chosen metric.

Heterogeneity, or equivalence, across research studies can cross out the option of conducting a meta-analysis; however, even when statistical groupings are reasonable, this remains a problem. Proper testing for heterogeneity is necessary, except when it is evident at a glance "that effects are consistent in magnitude and direction" (Polit & Beck, 2012, p. 662). Creating a forest plot will achieve a visual assessment of heterogeneity. The effect sizes of the studies will be estimated with the graph and jointly with a 95% confidence interval around the estimates (Polit & Beck, 2012).

A researcher conducting a meta-analysis frequently needs to consider how bias in outcome reporting could impact the effect size. Several methods can be undertaken to address bias. If the researcher is including studies that are randomized by group (e.g., family unit, school), he or she may need to use the intraclass correlation coefficient to examine whether observations within clusters are independent. Reporting (publication) bias may also impact the effect size. To address publication bias, researchers can

use a funnel plot to examine the distribution of effect sizes across studies included in the review. If there is no bias, the funnel plot should be symmetrical. If bias is found, researchers can use the trim and fill method to impute the values of studies that are assumed missing because of publication bias and recalculate the effect size (Duval, 2005). Variation of rigor in study design and inclusion of small studies in the meta-analysis may also lead to bias. Again, researchers can use funnel plots to examine this bias. They can also calculate the effect size with and without the small or less rigorously designed studies (Littell et al., 2008).

■ Conclusion

Critical appraisal of research is a fundamental part of evidence-based practice. It begins with understanding the research process in order to carefully and systematically evaluate studies to judge their relevance for clinical practice. To determine significance of the research you are considering, examine the following areas:

- Does the study test a stated hypothesis?
- Who is being studied? How were participants selected?
- Is the research design appropriate for the research question/hypothesis?
- Is each feature of the research design clear and replicatable?
- What measures were used and how were the data collected?
- What are the results of the study, and are they statistically significant?

This chapter summarized the different types of quantitative research to support critical appraisal of studies to improve patient outcomes.

Reflective Activities

1. How are variables operationalized?
2. Which variable—independent, dependent, or confounding—is the focus of the research study?
3. What key techniques separate experimental from nonexperimental research designs?
4. What research design would best compare two patient interventions (e.g., for lowering cholesterol)?
5. Why might a practitioner use a quasi-experimental research design in the practice setting?
6. How does a systematic review differ from a meta-analysis?

References

Babbie, E. (2010). *The practice of social research*. Belmont, CA: Wadsworth.

Brown, S. J. (2012). *Evidence-based nursing: The research-practice connection* (2nd ed.). Burlington, MA: Jones & Bartlett Learning.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago, IL: Rand McNally.

Durbin, C. G., Jr. (2004). How to come up with a good research question: Framing the hypothesis. *Respiratory Care, 49*(10), 1195–1198.

Duru, O. K., Sarkisian, C. A., Leng, M., & Mangione, C. M. (2010). Sisters in motion: A randomized controlled trial of a faith-based physical activity intervention. *Journal of the American Geriatrics Society, 58*, 1863–1869.

Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Bornstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 128–144). Chichester, England: Wiley.

Gibbs, L., & Gambrill, E. (1996). *Critical thinking for social workers: A workbook*. Thousand Oaks, CA: Pine Forge Press.

Grinnell, R. M., Jr., & Unrau, Y. A. (2011). *Social work research and evaluation: Foundations of evidence-based practice* (9th ed.). New York, NY: Oxford University Press.

Grinnell, R. M., Jr., Unrau, Y. A., & Williams, M. (2011). Group-level designs. In R. M. Grinnell, Jr., & Y. A. Unrau (Eds.), *Social work research and evaluation: Foundations of evidence-based practice* (9th ed., pp. 258–283). New York, NY: Oxford University Press.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0). Chichester, England: Wiley. Retrieved from http://www.cochrane-handbook.org/

Keele, R. (2011). *Nursing research and evidence-based practice*. Sudbury, MA: Jones & Bartlett Learning.

Kroeze, W., Oenema, A., Dagnelie, P. C., & Brug, J. (2008). Examining the minimal required elements of a computer-tailored intervention aimed at dietary fat reduction: Results of a randomized controlled dismantling study. *Health Education Research, 23*, 880–891.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York, NY: Oxford University Press.

Polit, D. F., & Beck, C. T. (2012). *Nursing research: Generating and assessing evidence for nursing practice* (9th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Rubin, A., & Babbie, E. (2011). *Research methods for social work*. Belmont, CA: Brooks/Cole Cengage.

Schmidt, N. A., & Brown, J. M. (2012). *Evidence-based practice for nurses: Appraisal and application of research* (2nd ed.). Burlington, MA: Jones & Bartlett Learning.

Shah, P. S., & Shah, J. (2010). Maternal exposure to domestic violence and pregnancy and birth outcomes: A systematic review and meta-analysis. *Journal of Women's Health, 19*, 2017–2031.

Solomon, P., & Draine, J. (2010). An overview of quantitative methods. In B. Thyer (Ed.), *The handbook of social work research methods* (2nd ed., pp. 26–36). Thousand Oaks, CA: Sage.

Surkan, P., Peterson, K., Hughes, M., & Gottlieb, B. (2006). The role of social networks and support in postpartum women's depression: A multiethnic urban sample. *Maternal and Child Health Journal, 10*(4), 375–383.

Yegidis, B. L., & Weinbach, R. W. (2009). *Research methods for social workers* (6th ed.). Boston, MA: Pearson.