Research Methodology: Designing a Research Study

Kenneth W. Schafermeyer and Peter D. Hurd

ABSTRACT: Using examples relevant to managed health care, this article describes ways of controlling variables in research, compares experimental versus nonexperimental research, illustrates several basic types of research designs, and describes various methods of sampling. This overview of some basic research concepts is designed to help managed care pharmacists critically evaluate primary research in areas such as drug efficacy trials, pharma-

coeconomics, and patient intervention programs. A second objective is to help managed care pharmacists frame research questions so that they can appropriately use data for decision making.

Key Words: Research methodology, Research design, Sampling, Managed care research

J Managed Care Pharm 1998; 504-514.

This is the second in a series of three articles that discuss research methodology as it relates to managed care pharmacy. The purpose of this series is to provide an overview of some basic research concepts so that managed care pharmacists can critically evaluate primary research in areas such as drug efficacy trials, pharmacoeconomics, and patient intervention programs. A second objective is to help managed care pharmacists frame research questions so that they can appropriately use data for decision making.

In the first article, Brenda Motheral described the concepts of hypothesis testing, measurement, reliability, and validity.¹ This, the second article, will focus on basic research design issues, and the third article will discuss basic statistical techniques.

Since this article builds on the concepts introduced in the previous article, it would be best to read the articles in sequence. To help the reader, we have repeated a few points in this article and have included a basic glossary of terminology introduced in the first article (Figure 1). This article will describe ways of controlling variables in research, compare experimental versus nonexperimental research, illustrate several different types of research designs, and describe various methods of sampling.

CONTROLLING VARIABLES IN RESEARCH

As Motheral described in the first article in this series, dependent and independent variables are important parts of research design. *Dependent variables* are the outcome measures of an experiment. They might include the cost of various therapies, the satisfaction of patients with their treatment, or the number of hospital visits for asthmatics. These measures are predicted to be dependent on the experimental conditions, or quasi-experimental situations, that have been monitored. For example, the cost of therapy will depend on the type of treatment program provided by the pharmacy benefit manager (PBM); similarly, patient satisfaction will be compared across different hospital settings. The *independent variables* are determined by the researcher prior to the study; thus the status of

AUTHORS

KENNETH W. SCHAFERMEYER, R.Ph., Ph.D., is Associate Professor and Director of Graduate Studies at the St. Louis College of Pharmacy, St. Louis, MO. PETER D. HURD, Ph.D., is Professor and Director of the Division of Administrative Sciences and Liberal Arts at the St. Louis College of Pharmacy, St. Louis, MO.

AUTHOR CORRESPONDENCE: Kenneth W. Schafermeyer, Ph.D., St. Louis College of Pharmacy, 4588 Parkview Place, St. Louis, MO 63110.

ACKNOWLEDGEMENT: The authors would like to thank Brenda R. Motheral, R.Ph., Ph.D., assistant professor at the University of Arizona College of Pharmacy, for her assistance with the manuscript.

CE CREDIT: This is article number 233-000-98-005-H04 in AMCP's continuing education program. It affords 2 hours (2 CEU) of credit. Learning objectives and test questions follow on page 515.

Copyright© 1998, Academy of Managed Care Pharmacy, Inc. All rights reserved.

Figure 1. Glossary

Following are some of the key terms introduced in the first article of this series, in order of appearance:

Hypothesis—a conjectural statement of the relation between two or more variables. A hypothesis must: 1) be declarative (i.e., in sentence form); 2) express a relationship between two or more variables; and 3) be capable of empirical testing.

Dependent variables—the outcome variables; the variables that are expected to vary depending upon the effect of the independent variables. The dependent variables are those that one is trying to explain.

Independent variables—those variables which are manipulated by the experimenter or those which are expected to affect the dependent variable.

Construct—a concept deliberately and consciously invented or adapted for a specific scientific purpose.

Operational definition—assigns meaning to a construct or a variable by specifying the activities or operations necessary to measure it.

Measurement—the assignment of numerals to objects or events according to rules. The rules used to assign numerals to objects define the type of measurement. Four levels of measurement exist: nominal, ordinal, interval, and ratio.

Nominal measurement—the lowest level of measurement, assigns numbers (or symbols) to objects for identity purposes rather than to give the connotation of rank order. Numbers used for sport teams' uniforms are nominal (e.g., number 10 is not considered higher or better than number 2).

Ordinal measurement—requires that objects in a set be rank ordered. Distances between attributes are not equal and have no meaning beyond indicating a more-or-less relationship between each category. That is, a given category of objects has more or less of an attribute than another category of that object. Ordinal measures do not have an absolute zero. As an example, temperature could be given the following values: 1=cold, 2=cool, 3=room temperature, 4=warm, 5=hot.

Interval measurement—requires numerically equal distances, the classic example being temperature (in Fahrenheit or Celsius). A change in temperature from 1 to 2 degrees Fahrenheit is the same distance as a change from 20 to 21 degrees. However, as with ordinal data, there is no absolute zero (i.e., "0" does not mean a lack of temperature).

Ratio measurement—requires a natural or absolute zero that has empirical meaning. A value of zero on the ratio scale indicates that the object has none of the property being measured. Accordingly, fractions or ratios are meaningful with a ratio variable.

Reliability—the similarity of results provided by independent but comparable measures of the same object. Synonyms for reliability are stability, dependability, and consistency. Another way of viewing reliability is by considering errors of measurement. While a certain amount of measurement error occurs each time a measurement is made, the goal is to minimize the amount of measurement error because the greater the measurement error, the more unreliable the instrument.

Construct validity—refers to the degree to which a variable accurately reflects the phenomenon which it purports to measure. Construct validity can be assessed by examination of convergent and discriminant measures.

Convergent measures—address whether the construct correlates with other concepts with which one would expect it to correlate.

Discriminant measures—involve examining correlations with measures from which a variable is suppose to differ.

Internal validity—refers to the approximate validity with which we infer that a relationship between two variables is causal. There are a number of threats to internal validity: 1) history; 2) maturation; 3) testing; 4) instrumentation; 5) statistical regression; 6) selection bias; and 7) experimental mortality.

External validity—the validity with which we can generalize to and across persons, settings, and times. Generally speaking, a lack of external validity represents an interaction between study variables and/or subjects and the environment.

Adapted from "Research Methodology: Hypotheses, Measurement, Reliability, and Validity," by Brenda Motheral (JMCP, July/August 1998).

independent variables have been determined prior to, and are independent of, the collection of data.

A well-designed research study will maximize the impact of the variables that really matter in developing a greater understanding of the subject under study, and minimize the influence of those factors that are not under study—analogue to separating the signal from the noise by using a good stereo receiver to reduce background noise. Sometimes researchers refer to this as maximizing *experimental variance* (that which the investigator has manipulated or is of special interest) and

minimizing or controlling extraneous variance that might dilute or mask the effects under study. To minimize the effects of *extraneous variance*, researchers have used a number of methods, including: holding variables constant; matching; building variables into the design; using statistical controls; and, most important, randomizing.

Hold Constant

In an attempt to control for factors that could lead to an erroneous interpretation of results, researchers will hold a

particular characteristic constant. Sometimes this is done through a protocol that excludes certain patients (e.g., no hypertension). In other cases, the decision is based on who will be included; for instance, only patients who have taken a given drug for at least six months might be considered in a particular study. Each criterion for entry into the study is designed to control factors that might lead to difficulty in interpreting the data. This tactic is a double-edged sword; holding some characteristics constant will decrease the "noise" in a study that might dilute the results, but also limits the extent to which the results can be generalized to groups with similar characteristics.

Matching

One way to control unwanted differences between research subjects is to select them so that different groups comprise very similar pairs. Studies of twins have compared the influences of various interventions, including the effects of varying environments when twins are raised in different homes. Some research designs attempt to match subjects by such demographic factors as age, gender, weight, education, and stage of an illness. While matching has advantages, the largest drawback often is the difficulty of finding people with similar characteristics and determining how close such similarities must be. Patients will not have identical ages; is five years older or younger an acceptable range? It is possible that another bias could be added to the study when one tries to match on only a few characteristics. For example, matching hypertensives' gender, age, and weight with those with normal blood pressure might inadvertently add a racial imbalance, because some ethnic groups are more prone to high blood pressure. Matching also adds a significant amount of clerical work in collecting information and making matches.

Build the Variable into the Design

Another way to control for factors that may have an impact on results is to make those components a part of the research design. For example, one might expect that use of mail-order pharmacy would affect patient satisfaction with pharmacy services. Making this variable part of the data collection (e.g., determine how often patients used mail-order pharmacy benefits during a year) or part of the patient selection process (e.g., half who have had mail-order claims in the past six months and half who have had only in-pharmacy claims) can allow the researcher to analyze the results for possible differences.

Statistical Controls

Occasionally, one can rely on statistical analysis to minimize bias. If age appears to be influencing satisfaction, and different groups have different mean ages, statistical techniques can help control for any potential bias that could explain the results. While using these techniques is sometimes required (e.g., some disease states are related to old age), it is best to design studies to minimize the need for statistical control of extraneous influences.

Randomization

By randomly distributing extraneous variables across study groups, researchers can reduce the potential for individual differences (such as age, gender, geographic location, health status, or educational level) to bias the study in some systematic way. Researchers often use randomization in the *assignment* of subjects to the independent variable conditions in an experiment. For example, subjects with asthma may be randomly assigned to either standard care or a high-powered education/intervention program. By randomly assigning individuals to one of these conditions, the experimenter hopes to control for any differences between groups that are not expressly controlled by the research design.

A very different way of using randomization is in *random selection* of subjects; a population of individuals is identified and research subjects are randomly selected from this grouping. The researcher who randomly chooses individuals in a benefits program to receive a survey is using this type of randomization. While survey research depends on randomization to acquire a representative sample of a population, one also could randomly select hospitals to survey or randomly select journals to evaluate for advertising content; thus, random selection can apply to a variety of situations.

Inexperienced researchers have a tendency to confuse random assignment and random selection. This type of confusion could lead a well-intentioned but misguided researcher to criticize a hospital experiment because patients were assigned to groups (e.g., half receive placebo and half receive active drug) rather than being randomly selected from the hospital population. This argument is somewhat like criticizing a study in which rats were randomly assigned to receive different doses of an experimental drug, because the researcher failed to randomly select from the population of rats. In both cases, randomization should avoid bias in the assignment of the subjects to the experimental or control groups. This spreads the effects of uncontrolled variables across the various experimental conditions. In this case, random assignment to groups is the appropriate concern, and random selection really is not an issue.

EXPERIMENTAL VERSUS NONEXPERIMENTAL

Once the objectives of a research study have been established, the next step is to formulate the research design. It is usually best for the researcher to write the objective in the form of a question. The type of question then determines whether an experimental or nonexperimental design is needed. If the question asks *why* something happened or whether one event caused another, then the researcher usually needs to conduct an experiment. (Potential limitations, however, may not always make this possible.) If, on the other hand, the question asks *what* or *how much*, then a nonexperimental design would be appropriate. (As discussed later in this article, it is sometimes convenient to categorize nonexperimental designs as either preexperimental or quasi-experimental.)

Experimental Research

The distinction between experimental and nonexperimental approaches to research is an important one. An experiment is a scientific analysis in which one or more independent variables is manipulated or controlled in order to determine the effect that this manipulation may have on the dependent variable(s).² For example, a researcher may manipulate the study variable by administering one of several drugs or dosages (independent variables) and observe the effect on blood pressure (dependent variable). Of course, other variables, such as age, gender, diet, or exercise also may affect blood pressure. To eliminate these factors from influencing the results of the study, the researcher must control these extraneous variables. In other words, variance should be minimized for the control variables but allowed for the study variables. By assuring that the effect being studied is actually the study variable, and not another factor, the researcher is conducting an experiment.

Nonexperimental Research

In nonexperimental research, the investigator cannot manipulate the study variable or randomly assign treatments or subjects. The researcher has very little control over the research environment and, in fact, often is working with data that were collected long before the analysis was conducted.³ Nonexperimental research lacks the scientific rigor necessary to draw conclusions about the effect of one variable on another. Nevertheless, nonexperimental research is common and does have value. Several examples of nonexperimental research are described below.

Descriptive Research

This type of investigation systematically describes the facts and characteristics of a given population or area of interest. It usually does not seek to explain relationships, test hypotheses, make predictions, or get at meanings and implications.⁴ Descriptive research typically reports sums, means, ranges, and variance or standard deviations. Descriptive research—probably the most common type of research conducted by managed care organizations (MCOs)—is used to report expenditures and utilization by group, patient, medical specialty, physician, drug class, drug, etc. This information can be used to provide comparisons among groups or during different time periods. It is important to note that descriptive research attempts to report *what* but not to answer the question *why*. For example, descriptive research would be used to answer the following questions: How many patients are taking a particular brand of ACE inhibitor? What is the average drug cost per patient per month (PMPM)? How do prescribing patterns for antidepressant prescriptions vary among medical specialties?

Case Study Research

This type of inquiry intensively studies the background, current status, and environmental interactions of a given group. Rather than looking at an entire population or a ran-

dom sample, a case study usually focuses on a single patient or on small groups. Since the focus is rather narrow, case studies do not provide valid information about the population as a whole; the results, therefore, cannot be generalized. Case studies do, however, provide useful anecdotes or examples to illustrate possible relationships. Follow-up research focusing on specific hypotheses and using proper sampling methods (discussed later in this article) would be needed to confirm the relationships.⁵ For example, a case study may ask "What is the effect of a nicotine patch on this patient?" or "How did this medical group respond to the new formulary?"

Correlational Research

This type of investigation studies the *extent* to which variations in one factor correspond with variations in one or more other factors. Examining the extent of a relationship often is superior to merely answering whether or not an effect is present or absent.⁶ An important distinction must be made here. Correlational research examines the relationship of two or more variables but does not determine the effect of one variable on another. For example, one may find that both ice cream sales and drownings increase in the summer. Obviously, just because these two events are associated does not mean that one caused the other. While this example is purposely absurd, inexperienced investigators reviewing a prescription claims database may find equally spurious relationships among variables and incorrectly conclude, for example, that closed formularies result in increased costs. While one may discover that variables are associated, a cause-and-effect relationship cannot be established without a design that effectively eliminates all rival hypotheses. For example, a study may question how age relates to diastolic blood pressure. While age is the variable being measured and compared, the researcher cannot conclude that age, by itself, is the cause of high blood pressure. As we know, age is related to certain physiological changes that, in turn, may explain the results.

Exploratory Research

Exploratory research looks for possible relationships among numerous variables. No specific relationships are hypothesized; the investigator merely is trying to determine whether any relationships exist that should be investigated more thoroughly through scientific experiment. For example, exploratory research could be used to determine whether unexpected side effects caused by a new drug are related to dosage, demographic factors, or other concurrent therapy. If a new drug were being used more frequently than anticipated, a researcher might want to explore further to determine whether usage is related to patient age or gender, whether there are geographic variations, which types of physicians are prescribing the drug, and what other drugs are being prescribed concurrently.

Qualitative Research

Not all research involves the analysis of numbers. Have you ever found that a quote by an individual conveys the findings

of a research project? Have you ever read through a series of written responses in order to determine categories for classifying these responses? Each of these examples has the look and feel of qualitative research.

Sometimes qualitative research is conducted by observers who interact with others. Schachter, for example, collected quotes and written descriptions from individuals who were living with epilepsy.⁷ Sometimes the observer is a part of the process. For example, perhaps a pharmacist is administering an asthma treatment program rather than collecting information from others involved in the program. The pharmacist describes the program, including some of the successes as well as some of the changes necessary to optimize outcomes. Written responses, interviews, and observations are included in the information that is collected, evaluated, synthesized, and reported to others. The pharmacy literature, as well as the managed care literature, does not report a great deal of truly qualitative research. The nursing profession, however, has more examples of this type of research.^{8,9} The combination of qualitative information with quantitative data can be a most effective way to promote knowledge.

Open-ended questions in patient satisfaction surveys that ask respondents to describe services that they like or dislike also constitute qualitative research. Pilot studies often include such open-ended questions to help investigators frame appropriate questions for larger studies.

Managed care organizations commonly conduct nonexperimental research to produce a wealth of valuable information. It is important to note, however, that such investigation is seldom driven by hypotheses and lacks the scientific rigor to examine cause-effect relationships. Since the control that forms the basis of scientific investigation is not rigorously applied in nonexperimental research, researchers should use caution to avoid drawing improper conclusions about the effect of one variable on another when conducting nonexperimental investigation.

LIMITATIONS OF EXPERIMENTAL RESEARCH

While well-designed experimental research is superior to nonexperimental research for establishing causal claims, it does have some disadvantages. Obviously, it takes more time and money to conduct experimental research. There is another important problem known as "generalizability." As an illustration, consider a typical clinical investigation for a potential new drug. The researchers identify a group of individuals who are relatively similar (e.g., young, healthy males) and randomly assign them to groups receiving either the drug (experimental group) or placebo (control group). To control extraneous variables, the subjects do not have comorbidities, dosage regimens are adhered to strictly, and, to the extent possible, other variables, such as diet and exercise, are controlled, or at least measured as possible independent variables. The experiment also may be blinded—neither the person making the measurements nor the research subject knows who is getting the drug

and who is getting the placebo. In some cases, the investigator may use a crossover design, which involves switching research subjects from one dosage form to the other after a designated time, so that individuals can serve as their own controls.

The results of this rigorously controlled experiment provide evidence of the effectiveness of the drug *under ideal circumstances*. Of course, once a drug is released on the market, these ideal circumstances no longer exist; physicians vary in their prescribing patterns, and patients vary in age, gender, health status, comorbidities, and degrees of compliance. The actual performance of a drug on the market, where many extraneous variables exist, often is different from the results found in the controlled clinical investigation. Therefore, while true experimental research is scientific, it is not always completely generalizable to actual patient care settings.

RESEARCH DESIGN

There are many ways to design a research study; the various methods have been categorized as preexperimental, experimental, and quasi-experimental.¹⁰ While many research methods texts describe experimental designs as superior, and consider preexperimental and sometimes even quasi-experimental designs as inadequate, it must be recognized that each design can be valuable when used for the purpose for which it is intended. One must be aware of the limitations of each design and avoid drawing unwarranted conclusions that are not supported by the design.

Following is a description of a few of the more common research designs, including a description of their uses and limitations in terms of threats to internal and external validity. Figure 2 defines the various threats to internal and external validity, many of which were described in the previous issue of JMCP.¹¹ To aid in the discussion, each research design is given a name and is illustrated by a symbolic notation originally developed by Campbell and Stanley.¹² An X represents the exposure of a group to an intervention or treatment variable whose effects are to be measured; an O refers to an observation or measurement. The Xs and Os in a given row are applied to the same persons or group and are listed from left to right in order of occurrence; Xs and Os vertical to one another occur to different groups simultaneously. An R designates that subjects are randomly assigned to the experimental and control groups. Figure 3 summarizes the discussion for each design.

Preexperimental Research Designs

In some cases it is impossible to measure a dependent variable both before and after an intervention. Likewise, it may not be possible to establish a control group. In those cases in which a true experiment cannot be conducted, a preexperimental design may be useful for exploring data or developing ideas. These research designs, however, do not have sufficient internal validity to establish cause-and-effect relationships.

Figure 2. Factors Jeopardizing Internal and External Validity

Internal Validity

1. *History*—the specific events occurring between measurements that can account for the findings.
2. *Maturation*—processes within the respondents which change with the passage of time, including growing older, growing hungrier, growing more tired, and the like.
3. *Testing effect*—the effect that a pretest may have on the score of a post-test.
4. *Instrumentation*—changes in the measuring instrument between the pretest and post-test or changes in the observers or scorers used which may produce changes in the obtained measurements.
5. *Statistical regression* (or regression toward the mean)—movement toward a mean that occurs in measurements when individuals are assigned to a group based on their scores. Individuals with very high or very low scores are more likely to score closer to the mean on subsequent measurements.
6. *Selection bias*—differential selection of respondents for the comparison groups.
7. *Experimental mortality*—differential loss of respondents from the comparison groups.
8. *Selection-maturation interaction, etc.*—in certain multiple-group quasi-experimental designs, such as time-series analysis, is confounded with (i.e., might be mistaken for) the effect of the experimental variable.

External Validity

9. The *reactive* or *interaction effect* of testing, in which a pretest might increase or decrease the respondent's sensitivity or responsiveness to the experimental variable and thus make the results obtained for a pretested population unrepresentative of the effects of the experimental variable for the unpretested universe from which the experimental respondents were selected.
10. The *interaction effects* of selection biases and the *experimental variable*.
11. *Reactive effects of experimental arrangements*, which would preclude generalization about the effect of the experimental variable upon persons being exposed to it in nonexperimental settings.
12. *Multiple-treatment interference*, likely to occur whenever multiple treatments are applied to the same respondents, because the effects of prior treatments are not usually erasable.

Note: Although points 8,10,11, and 12 are not detailed in the articles in this series, they are described here to provide a comprehensive list.

Adapted from: "Experimental and Quasi-Experimental Designs for Research," by DT Campbell and JC Stanley, (Boston: Houghton Mifflin, 1963).

Design 1—The One-Shot Case Study

X O

First, consider an example of this type: A PBM sends educational information to physicians to discourage inappropriate prescribing of a given drug (intervention X). A few months later the PBM measures utilization of the drug (observation O). This design has a complete absence of control and no provision for comparison prior to the intervention, other than an impression or intuition. Accordingly, it is unknown how much utilization changed from pre- to post-intervention or how this change would compare to a control group of physicians who were not subject to the educational information. Despite the fact that this design has no internal validity, inexperienced researchers often misuse information gleaned from such studies as a basis for program changes or, more commonly, to assess changes that already were implemented. This design should be used solely for exploratory purposes.

Design 2—The One-Group Pretest-Posttest Design

O₁ X O₂

The previous example has been modified here to measure utilization both before (O₁) and after (O₂) conducting the

physician education program (X). This method is better than Design 1 because using the same subjects for O₁ and O₂ avoids threats to internal validity due to selection bias and experimental mortality. However, there is still no control group and, consequently, no assurance that the X was the only or even the major factor resulting in a difference between O₁ and O₂. There are a number of plausible rival hypotheses:

- ▲ *History*—New drugs could have come on the market between O₁ and O₂, new side effects may have been discovered, or there may have been changes in the patients enrolled in the managed care organization.
- ▲ *Maturation*—Patients are getting older (especially if only continuously enrolled patients are studied), or experience has made patients less enthusiastic about taking the drug.
- ▲ *Testing effects*—The experience of taking a pretest could sensitize the subject to the intervention being measured and affect the results of the posttest. (This would be an issue when surveying patients, but not when using a database.)
- ▲ *Instrumentation*—Changes in the test or the observers making the measurements could create a difference between O₁ and O₂. (Again, this would be an issue for survey research but not for analysis of an existing database.)

Figure 3. Sources of Threats to Internal and External Validity

Internal									External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Pre-Experimental Designs:												
1. One-Shot Case Study X O	-	-				-	-			-		
2. One-Group Pretest-Post-test Design O ₁ X O ₂	-	-	-	-	?	+	+	-	-	-	?	
3. Static-Group Comparison X O O	+	?	+	+	+	-	-	-		-		
True Experimental Designs:												
4. Pretest-Post-test Control Group Design R O ₁ X O ₂ R O ₁ O ₂	+	+	+	+	+	+	+	+	-	?	?	
5. Post-test-Only Control Group Design R X O R O	+	+	+	+	+	+	+	+	+	?	?	
Quasi-Experimental Designs:												
6. Time Series O ₁ O ₂ O ₃ O ₄ X ₅ O ₆ O ₇ O ₈ O ₉	-	+	+	?	+	+	+	+	-	?	?	
7. Nonequivalent Control Group Design O ₁ X O ₂ O ₁ O ₂	+	+	+	+	?	?	+	-	-	?	?	

Note: In the table, a minus indicates a definite weakness, a plus indicates that the factor is controlled, a question mark indicates a possible source of concern, and a blank indicates that the factor is not relevant for the particular design.

Source: "Experimental and Quasi-Experimental Designs for Research," by DT Campbell and JC Stanley, (Boston: Houghton Mifflin, 1963).

▲ Regression toward the mean—If high prescribers in O₁ are studied again in O₂, their utilization is much more likely to decrease (i.e., regress toward the mean) than it is to increase or even stay the same.

Design 3—The Static-Group Comparison

X O
O

An example of this design is a comparison of PMPM prescription costs for MCOs that maintain restrictive formularies with those of MCOs that have open formularies. Using this

design, it would be impossible to defend a conclusion, for example, that open formularies reduce program costs. In this case there is a control group that did not receive the intervention but no baseline measurement to determine whether the dependent variable changed after the intervention. While this form does reduce the problems related to history, testing effects, instrumentation, and statistical regression (and to a lesser extent maturation), it still presents serious internal validity problems: ▲ Selection bias—The two groups being studied may be quite different. Their demographics, utilization patterns, and the drugs covered are unlikely to be identical, the prescribers may

be different, and there may be regional differences in prescribing patterns.

▲ **Experimental mortality**—This design assumes that the two groups were essentially the same before the intervention. Even if the demographics of the employer groups had been the same at one time, different drop-out rates for the groups are likely to cause significant changes over time. In this case, groups that experienced rapidly increasing drug expenditures would be more likely to adopt restrictive formularies. Therefore, those groups with high expenditures are more likely to move out of the open formulary group and into the closed formulary group, thereby affecting the posttest measurement.

Experimental Research Designs

Well-designed experiments have strong internal validity. While bench lab scientists and clinical investigators can control variables and randomly assign subjects to treatment groups, this usually is not possible for investigators working with prescription claims databases. MCOs, therefore, cannot usually conduct true experimental research. Still, it is important for any researcher to understand the common research design methods for experiments in order to select the best alternative research design, fully understand the potential threats to validity, and recognize the potential shortcomings of his or her research. In research, investigators always deal with probability, not certainty. A researcher never proves anything; but only finds evidence to support or refute a hypothesis. When working outside of the true experimental designs, researchers accept that there is less certainty (i.e., lower probability) that the results they find are valid. However, research carries a greater probability of being correct than do intuition or past experience.

Design 4—Pretest-Posttest Control Group Design

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

In this case, subjects are randomly assigned to either the experimental group (top) or the control group (bottom). To determine that a treatment (X) had an effect, the pretest measurements (O_1) will be similar, while the posttest measurements (O_2) will be different for the experimental and control groups. Many clinical investigations use this type of design. There are several variations of this method, such as double-blind and crossover studies previously mentioned.

The lack of generalizability is a potential problem with Design 4. Consider either a quality-of-life study or a patient satisfaction survey in which patients are given pretests. The posttest results may be affected by the fact that the patients have been sensitized by the pretest and now start to take more notice of the quality of services they receive or symptoms they may have ignored previously. This effect is known as an "interaction of testing and X"—a threat to external validity. (This problem would not be as common in more objective measurements such as blood pressure.) The following design offers a control over this threat to external validity.

Design 5—Posttest-Only Control Group Design

$$\begin{array}{ccc} R & X & O \\ R & & O \end{array}$$

To many research workers, this design may look insufficient because there is no pretest. Although the pretest is a concept deeply imbedded in the thinking of research workers, it is not essential to true experimental designs.¹³ For some, giving up the pretest means not knowing for sure whether X affected the O. However, the essential element that makes this research design acceptable (and different from Design 3) is randomization, which is by far the most adequate means of avoiding initial biases between groups. This design is as good as Design 4 and for some studies it is even better. By eliminating the pretest, the researcher eliminates the possibility of interaction between the pretest and X and, consequently, removes one threat to external validity. This is especially useful when making subjective measurements, such as patient satisfaction and quality of life, that can be affected by prior testing of the research subjects.

Quasi-Experimental Research Designs

Quasi-experimental research designs have a very important role in research, especially when the use of human subjects makes a strict experimental design impossible, unethical, or impractical. As the term "quasi-experimental" implies, these designs appear much like a true experiment, except in the lack of control and randomization normally found in experimental designs. For example, some quasi-experimental projects (such as Design 7) have nonequivalent control groups. Rather than randomizing subjects to treatment conditions, a study may compare hospitals, one with a new treatment and one with the existing program. Since there is no control for the type of patient at the two hospitals and for the level of care at each hospital, biases may creep into the results.¹⁴ In other quasi-experimental designs, such as the time-series analysis (Design 6), there may not be a control group. Since it is often impossible for MCOs to employ experimental designs, quasi-experimental designs often are the best alternative.

Design 6—Time-Series Analysis

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

This design resembles Design 2 in lacking a control group, using before-and-after measures, and not including randomization. However, this design is superior because the repeated measurements taken over a period of time help reduce some of the internal validity problems inherent in Design 2. If, for example, there is no appreciable difference in the first four observations, then a difference between O_4 and O_5 would not be due to maturation, testing, regression, or to a lesser extent, instrumentation. As in Design 2, selection bias and experimental mortality also are controlled.¹⁵

The chief threat to internal validity is history. For instance, changes in program design, patient enrollment, prescribers,

drug costs, or other factors that occur between O_1 and O_2 could explain an observed difference in drug utilization or expenditures. As with Design 2, a time-series analysis applied to survey research also is subject to the threat to external validity known as interaction of testing and X. (This problem, of course, does not occur when analyzing an existing database.)

Design 7—Nonequivalent Control Group Design

O_1 X O_2
 O_1 O_2

This is the same as Design 4 except that experimental subjects are not assigned randomly to the experimental and control groups. Certainly, even a nonequivalent control group is far superior to a design with no control group at all (e.g., Design 2). Although selection bias may be an issue in some cases, the more similar the experimental and control groups are in their recruitment, and the more this similarity is confirmed by the scores on the pretest, the more effective this control becomes.¹⁶ The control group eliminates history as a possible threat to internal validity. Statistical regression toward the mean still may be a problem, particularly when researchers try to establish control by matching subjects in the experimental and control groups. This design has the same problem with external validity as Design 6.

Multiple Treatment Designs

All of the previous designs assumed that there was a single treatment group, sometimes with a control group. Researchers often want to investigate the effects of two or more independent variables simultaneously. To illustrate this, consider a simple variation to Design 4 in which two or more treatment groups are simultaneously compared (e.g., different drugs or two different strengths) with a single control group.

Design 4A—Variation of Pretest-Posttest Control Group Design

Treatment Group 1 R O_1 X O_2
 Treatment Group 2 R O_1 X O_2
 Control Group R O_1 O_2

While this design helps to control the threats to internal validity, the investigator must be careful to assure that the experimental controls are the same for each group (e.g., the same instructions, measurement procedure, time frame, etc.).

Another example is a variation of Design 5 which divides both the treatment and control groups into two groups, one with an additional drug or vitamin supplement, for example, and one without. Therefore, this "2x2" design has four groups: one with both drugs, one with drug A, one with drug B, and a control group with neither drug. This design allows the study of each drug individually, plus the interaction of the two (see Figure 4).

Design 5A—Variation of Posttest-Only Control Group Design (2x2 Matrix)

Of course, there are many permutations to this design—2x3, 3x3, 2x2x2, etc. All include a control group and allow

Figure 4. 2x2 Matrix

		Drug A	
		Yes	No
Drug B	Yes	AB	B
	No	A	Control Group

study of interactions of two or more variables. Although only the 2x2 example is given here, one can readily see that multiple treatments can be applied to many different types of designs.

By understanding the advantages and potential limitations of various research designs, managed care pharmacists can more effectively critique the value of research described in the literature. Likewise, managed care pharmacists facing research questions should select the research design that most closely meets their needs. It is seldom possible to conduct a flawless research project; however, knowing the potential limitations of a given research project and minimizing them to the extent possible through proper research design and control helps reduce the possibility of drawing incorrect conclusions.

Longitudinal Versus Cross-Sectional Designs

Many of the studies conducted in managed care cover a particular time, either the same day or month, or a relatively short period of time. In an attempt to control for age differences, some projects use age as an independent variable, essentially placing subjects in various age groups. This type of research design sometimes is called cross sectional because one takes a slice of data from a particular time to compare the impact on age. A cross-sectional design can be compared to a longitudinal approach in which the same individuals are followed across time, sometimes for decades. The Baltimore longitudinal aging study is one of many that has continued to evaluate participants as they grow older.

Both approaches have strengths and weaknesses. The cross-sectional study is much less expensive than the longitudinal study, which requires repeated follow-ups, record keeping, and much more time to complete. The longitudinal study suffers from such threats to internal validity as the testing effect, instrumentation, and maturation. Cross-sectional designs, however, can lead to inaccurate inferences about age effects, because the younger age groups have experienced a different life than the older age groups. A cross-sectional study of intelligence scores might show that older groups score lower than younger groups. This would not, however, take into account the fact that older participants are much less likely to have finished college and have had less experience with multiple-choice tests. The researcher using a cross-sectional design, then, should be very cautious about making conclusions that may not be supported in longitudinal studies.

SAMPLING

The terms *population* and *sample* have special meanings in statistics. Population refers to the entire group that has been defined; sample refers to individuals who have been selected from that population. In some research, the entire population can be included in the study, as when all hospital employees receive a questionnaire or all patients enrolled in a plan are included in a drug utilization study. Other types of research are based on a sample of the entire population. The following paragraphs describe several ways in which a sample can be selected from a population.

Simple Random Sampling

This type of sampling often is used to randomly select study subjects from a population. The investigator assigns every individual a unique number and then turns to a random number table (or computer-generated random numbers) and selects the individuals whose numbers are generated. For a sample of 30, one could take the first 30 people whose numbers matched; if a random number were generated twice, or a number exceeded one's population, one would simply move on to the next viable number. To achieve simple random sampling, researchers need to generate what is called a *sampling frame*—a list of the people in the population. The key to this approach is that everyone has the same chance of being selected as a part of the research study. When people lack a unique number in a large population, such as a survey of residents in a large city, simple random sampling is difficult to conduct.

For example, consider how a researcher would use the yellow pages to design a simple random sample of the hospitals in a given community. Unless the researcher lives in a very large community, he or she might assign each hospital a number (the easiest way would be to start with 1 and continue on to the last hospital). If the community had 55 listings, for example, and the researcher wanted to select 20 hospitals, he or she would go to a series of random numbers, choose a random starting point, and select the first 20 two-digit numbers that are between 01 and 55. The researcher would use these randomly generated numbers to select hospitals from the telephone list (selecting 05 from the random number table means that the fifth hospital in the telephone listing would be included in the sample). When actually using a telephone book, the researcher would find it necessary to decide first whether any particular hospital listings should be excluded from the sample (e.g., psychiatric hospitals, multiple listings for large hospital complexes, phone numbers listed in advertisements, out-of-town listings, etc.). The objective and design of the study would help the investigator answer these questions.

Stratified Random Sampling

Sometimes the nature of the population or the research question being addressed requires a mix of responses. If, for example, a researcher wanted to compare small, medium, and

large hospitals, then he or she would need to select the sample so that it includes enough large hospitals. In the U.S., a simple random sample of hospitals would yield too many small hospitals at the expense of very few large hospitals for comparison, because there are relatively few really large hospitals compared to the medium and small hospitals.

One way to solve this problem is to modify simple random sampling by adding a set of decision rules that will assure a certain number of individuals for each group. In the hospital example, one might decide to take the first 30 hospitals randomly identified in the large, medium, and small hospital categories. In this case, one might have identified 30 small hospitals at random long before randomly identifying 30 large hospitals. This approach, however, assures the sample will include members of the groups necessary for comparison.

Systematic Sampling

With the increasing use of computers, systematic sampling has become very simple and more common. In this case, an investigator might decide to select every fifth (or every 500th) new prescription to get some sense of the mix of products used in a certain month in a specific benefit plan. Systematic sampling works as long as there is no systematic bias to affect the results. Claim forms may well be submitted in a random fashion; however, if one pharmacy sent in many months' worth of claims all at the same time, a systematic sample made at that time would over-sample that pharmacy's claims, which could result in an inaccurate picture.

Sometimes databases are sorted on some factor such as identification number, which usually has no systematic bias. Babbie,¹⁷ however, describes an example of periodicity that led to an inaccurate sample. During World War II, unit rosters were used to select soldiers for a study; every 10th member on each list was selected. Since sergeants were listed first on every list, they were never selected, leading to a biased sample. Researchers working in a managed care setting should have a good understanding of the database's organization and a cooperative working relationship with the computer people who might be able to avoid selection problems.

Sampling for Large National Groups

Sampling actually becomes very complicated when one tries to make judgements about a nation's political preferences based on a sample of approximately 1,000 people. One technique used to correct this problem involves determining multiple stages in the sampling process, attempting to omit bias in the selection process for each stage. To sample all the individuals listed in a city's phone book, the stages might include pages, columns, and rows. The researcher would randomly choose enough pages to complete the necessary sample. Then the researcher would randomly choose a column on each page and a phone number in each column.

This same technique could be modified for national samples, including appropriate stages for the research question,

such as region of the country, state, and city. Working with a company that does national sampling is often a simple way to assure reliable information. Sampling frames are seldom perfect; telephone numbers, for example, can be unlisted, and a few individuals do not have telephones while other families may have two or more.

Other Sampling Methods

The goals of the research project will help determine the sampling method that is most appropriate. Sometimes random sampling is not necessary. For example, in *convenience sampling*, the researcher uses subjects that are relatively easy to contact. Although unscientific, this simple method would be appropriate for a pilot test of a survey or for exploratory research. A convenience sample of the first 50 entries in a database can be selected to give a researcher a feel for the data that will be analyzed.

In some types of research, the most appropriate approach might be *snowball sampling*, in which a researcher begins with a few appropriate individuals and then asks them to help identify others with the characteristics appropriate for the study. When research has a specific purpose, the researcher selectively chooses only those individuals who are appropriate for their research question. If the investigator wants to know what successful PBM chief executive officers think about quality of care, he or she might contact these individuals and seek their consent for an interview.

CONCLUSION

Pharmacists practicing in managed care settings are increasingly faced with questions that require them to use basic research skills. By understanding basic research concepts, managed care pharmacists will be more effective in evaluating existing literature and using it appropriately. Pharmacists with

research skills can also improve the quality of information generated for reports and for making decisions that will help MCOs truly manage the prescription benefit.

▲ References

1. Motheral BR. Research methodology: hypotheses, measurement, reliability, and validity. *JMCP* 1998; 4: 382-88.
2. Kerlinger FN. Foundations of behavioral research. New York: Holt, Rinehart, and Winston, Inc., 1986: 348-49.
3. Bailey KD. Methods of social research. New York: The Free Press, 1987: 217.
4. Isaac S, Michael WB. Handbook in research and evaluation. San Diego: Edits Publishers, 1985: 46.
5. Isaac S, Michael WB. Handbook in research and evaluation. San Diego: Edits Publishers, 1985: 48.
6. Isaac S, Michael WB. Handbook in research and evaluation. San Diego: Edits Publishers, 1985: 49.
7. Schachter SC. Brainstorms: epilepsy in our words—personal accounts of living with seizures. New York: Raven Press, 1993.
8. Polit DF, Hungler BP. Essentials of nursing research: methods, appraisal, and utilization, 4th Edition. Philadelphia: Lippincott, 1997.
9. Burns N, Grove SK. The practice of nursing research: conduct, critique, and utilization. Philadelphia: W.B. Sanders, 1987.
10. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company, 1963.
11. Motheral BR. Research methodology: hypotheses, measurement, reliability, and validity. *JMCP* 1998; 4: 382-88.
12. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company, 1963: 6.
13. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company, 1963: 25.
14. Hurd PD, et al. Research design and data collection methods in Eucharist NO (Ed.), Health research design and methodology. Boca Raton, FL: CRC Press, 1990: 99-100.
15. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company, 1963: 37-43.
16. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company, 1963: 48.
17. Babbie E. The practice of social research, seventh edition. Belmont, CA: Wadsworth, 1995: 209-10.

Practitioner Update—Educational Objectives

Upon completion of this article, the successful participant should be able to:

1. differentiate between experimental and nonexperimental research.
2. describe various types of nonexperimental research and how they are used.
3. explain how to control the effects of extraneous variables.
4. describe the limitations of experimental research.
5. describe the various research designs and their strengths and weaknesses.
6. differentiate between longitudinal and cross-sectional research designs.
7. describe how the various methods of sampling from a population are used in research.

SELF-ASSESSMENT QUESTIONS

Use the following scenario to answer Questions 1-4.

In an attempt to assess the effectiveness of physician intervention letters for antiulcer medication prescribing, Medicaid prescription claims were analyzed for patients receiving this therapy. Half of these patients' physicians received an intervention letter. Those who would receive a letter were chosen at random. The prescribing patterns of the physicians receiving the intervention letter were compared with the prescribing patterns of those physicians who had not been sent an intervention letter. Results indicated that physicians receiving the letter used less-duplicative antiulcer medications.¹

1. The two groups (those that received or did not receive the letter) represent
 - a. the dependent variable.
 - b. the independent variable.
 - c. external validity.
 - d. random selection.
 - e. stratified random sampling.
2. The determination of which physician would receive a letter and which would

not was random. This would be an example of

- a. a dependent variable.
 - b. systematic sampling.
 - c. external validity.
 - d. random selection.
 - e. random assignment.
3. This design would best be described as
 - a. descriptive research.
 - b. case study research.
 - c. qualitative research.
 - d. experimental research.
 - e. quasi-experimental research.
 4. The research design used for this study is best described as a
 - a. one-group pretest-posttest design (Design 2).
 - b. static-group comparison (Design 3).
 - c. pretest-posttest control group design (Design 4).
 - d. posttest-only control group design (Design 5).
 - e. time-series analysis (Design 6).

5. When an organization conducts a survey based on a representative sample of U.S. managed care users, they most likely use

- a. simple random sampling.
 - b. stratified random sampling.
 - c. systematic sampling.
 - d. convenience sampling.
 - e. snowball sampling.
6. When a researcher pilot tests a questionnaire on his friends at work who are pharmacists, this probably would be
 - a. simple random sampling.
 - b. stratified random sampling.
 - c. systematic sampling.
 - d. convenience sampling.
 - e. snowball sampling.
 7. The type of research that describes narrative responses to open-ended questions is
 - a. descriptive research.
 - b. case study research.
 - c. exploratory research.

d. qualitative research.

e. quasi-experimental research.

8. Which of the following statements is most accurate regarding experimental research?

- a. Experimental research usually lacks internal validity.
- b. MCOs usually find experimental research easy to conduct because large databases provide ready access to control groups and comparison groups.
- c. Experimental research may have some weaknesses with regard to external validity because results determined through experimental conditions may not be generalizable to the actual practice environment.
- d. Experimental research is the only type of research that should be conducted by MCOs.
- e. Experimental research answers the question "what;" it does not answer "why."

9. Factors that dilute or mask the effects under study are best described as

- a. independent variables.
- b. dependent variables.
- c. experimental variables.
- d. quasi-experimental variables.
- e. extraneous variables.

10. Which of the following is true regarding longitudinal and cross-sectional designs?

- a. Longitudinal designs are much less expensive to conduct than cross-sectional designs.
- b. Cross-sectional designs eliminate the effect of age as an explanation for differences among groups.
- c. Longitudinal designs follow the same patients across time.
- d. Cross-sectional designs suffer from two threats to internal validity: the testing effect and maturation.

See text of article beginning on page 504 of this issue of *JMCP*.

This article qualifies for 2 hours of continuing pharmaceutical education (.2 CEU). The Academy of Managed Care Pharmacy is approved by the American Council on Pharmaceutical Education as a provider of continuing pharmaceutical education. This is program number 233-000-98-005-H04 in AMCP's educational offerings.

CE EXAM

DEMOGRAPHIC INFORMATION (not for scoring)

11. In what type of setting do you work (leave blank if none of the responses below applies)?

- a. HMO.
- b. PPO.
- c. Indemnity insurance.
- d. Pharmacy benefits management.
- e. Other.

12. Did this program achieve its educational objectives?

- a. Yes.
- b. No.

13. How many minutes did it take you to complete this program, including the quiz (fill in on answer sheet)?

14. Did this program provide insights relevant or practical for you or your work?

- a. Yes.
- b. No.

15. Please rate the quality of this CE article.

- a. Excellent.
- b. Good.
- c. Fair.
- d. Poor.

▲ Reference

1. Rascati KL, Okano GJ, Burch C. Evaluation of physician intervention letters, Medical Care 1995; 34: 760-66.

INSTRUCTIONS

This quiz affords 2 hours (.2 CEU) of continuing pharmaceutical education in all states that recognize the American Council on Pharmaceutical Education. To receive credit, you must score at least 70% of your quiz answers correctly. To record an answer, darken the appropriate block below. Mail your completed answer sheet to: Academy of Managed Care Pharmacy, 100N. Pitt Street, Suite 400, Alexandria, VA 22314. Assuming a score of 70% or more, a certificate of achievement will be mailed to you within 30 days. If you fail to achieve 70% on your first try, you will be allowed only one retake. The ACPE Provider Number for this lesson is 233-000-98-005-H04. This offer of continuing education credits expires October 31, 1999.

	A	B	C	D	E
1.	☐	☐	☐	☐	☐
2.	☐	☐	☐	☐	☐
3.	☐	☐	☐	☐	☐
4.	☐	☐	☐	☐	☐
5.	☐	☐	☐	☐	☐

	A	B	C	D	E
6.	☐	☐	☐	☐	☐
7.	☐	☐	☐	☐	☐
8.	☐	☐	☐	☐	☐
9.	☐	☐	☐	☐	☐
10.	☐	☐	☐	☐	☐

11. ☐ A ☐ B ☐ C ☐ D ☐ E

12. ☐ Yes ☐ No

13. Minutes _____

14. ☐ Yes ☐ No

15. ☐ A ☐ B ☐ C ☐ D

Participant Identification: Please type or print

Social Security # _____
For Identification Purposes Only

Date _____

Name _____
Last First Middle

Work Phone # _____

Company _____

Address _____
Street (with Apt. No.) or P.O. Box City State Zip

State and Lic. No. _____
State No.

Member Type: ☐ Active ☐ Supporting Associate
☐ Student ☐ Nonmember

Signature _____