Research Methodology: Hypotheses, Measurement, Reliability, and Validity

Brenda R. Motheral

ABSTRACT: Using real-world examples, this article examines hypothesis development, measurement, reliability, and validity. This article provides guidance for critically evaluating the published literature. In the changing health care industry, knowledge of research methodology will benefit pharmacists practicing in a managed care environment.

Key Words: Research methodology, Validity, Reliability, Hypotheses

J Managed Care Pharm 1998: 382-390

In the rapidly evolving health care industry, pharmacists are assuming a wide variety of roles, many of which were unimaginable even 10 years ago. Of all the skills needed by managed care pharmacists, education and training in research methodology is near the top of the list. Pharmacists involved in policy decisions regarding formulary coverage, patient education, and compliance programs would certainly benefit from a working knowledge of research methodology. To guide real-world decisions and program development, pharmacists must be able to critically evaluate primary research in drug efficacy trials, pharmacoeconomics, and patient intervention programs. Furthermore, pharmacists are increasingly conducting research in the managed care setting. Against this backdrop, *JMCP* is publishing a series of three articles that will discuss research methodology as it relates to managed care pharmacy. This first article focuses on the concepts of hypothesis testing, measurement, reliability, and validity. Study design will be covered in the second article, and the final article will discuss basic statistical techniques.

HYPOTHESIS TESTING

Science is not just a fact-gathering activity. Rather, scientists must have some guiding idea in order to know which facts to gather. This preconceived idea takes the form of a hypothesis, which is "a conjectural statement of the relation between two or more variables."[1] 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.[1] In other words, the variables of interest must be measurable.

An example is: "Patients who have a pharmacy benefit are more satisfied with their health insurance plan than patients who do not have a pharmacy benefit." In this example, the hypothesis is declarative in nature, takes into account two variables (i.e., provision of a pharmacy benefit and satisfaction), and is capable of empirical testing. However, consider the following hypothesis: "Pharmacists provide high-quality medication

▲ *AUTHOR*

BRENDA R. MOTHERAL, R.PH., M.B.A., PH.D., is currently Assistant Professor, University of Arizona, College of Pharmacy, Tucson, AZ. At the time this article was written, Dr. Motheral was Assistant Professor at St. Louis College of Pharmacy and Outcomes Researcher at Express Scripts, Inc.

AUTHOR CORRESPONDENCE: Brenda R. Motheral, Ph.D., College of Pharmacy, University of Arizona, P.O. Box 210207, Tucson, AZ 85721-0207.

CE CREDIT: This is article number 233-000-98-004-H04 in AMCP's continuing education program. It affords 2 hours (.2 CEU) of credit. Learning objectives and test questions follow on page 389.

counseling." This hypothesis does not express a relationship between two or more variables; it merely describes a characteristic of pharmacists. Examine another hypothesis: "Patients who have guardian angels are less likely to die in a motor vehicle accident than patients who do not have guardian angels." While the illustration is a little extreme, it provides a clear example of a hypothesis that is not amenable to empirical testing.

In addition to the requirements listed above, a good hypothesis is value free and avoids vague generalizations.[1] The terms "better" and "ought to" usually indicate a hypothesis that is normative (not value free) in nature, such as: "Pharmacists are better patient counselors than other health professionals." This statement reflects a value judgment that is difficult to empirically test in a meaningful way. The hypothesis "Closed formularies have a negative impact on patients" is fairly vague. A more specific hypothesis would be: "Closed formularies increase overall medical expenditures."

In published articles, a hypothesis is not always stated directly. The reader must infer, typically through examination of the stated research question, the actual hypothesis. Once the hypothesis has been identified, the reader can determine whether the hypothesis follows the requirements and guidelines for a good hypothesis. However, not all studies require hypotheses. While studies that examine a possible cause–effect relationship should have hypotheses, descriptive studies, such as those that describe rates of steroid inhaler use or calculate the cost of treating diabetes, do not require them.

CONSTRUCTS, VARIABLES, AND OPERATIONAL DEFINITIONS

Generally, hypotheses have independent and dependent variables. Dependent variables are those variables that one is trying to explain. Independent variables are those that are manipulated by the experimenter or are expected to affect the dependent variable. For example, a researcher may examine the influence of using a mail-order pharmacy on drug expenditures. In this case, use of a mail-order pharmacy is the independent variable, and drug expenditures represent the dependent variable. Variables may be independent or dependent, but not in the same study. For example, one study may examine patient characteristics that predict compliance, while another study may examine the effect of compliance on hospitalization rates. In the first case, compliance is a dependent variable; in the second study, compliance is an independent variable.

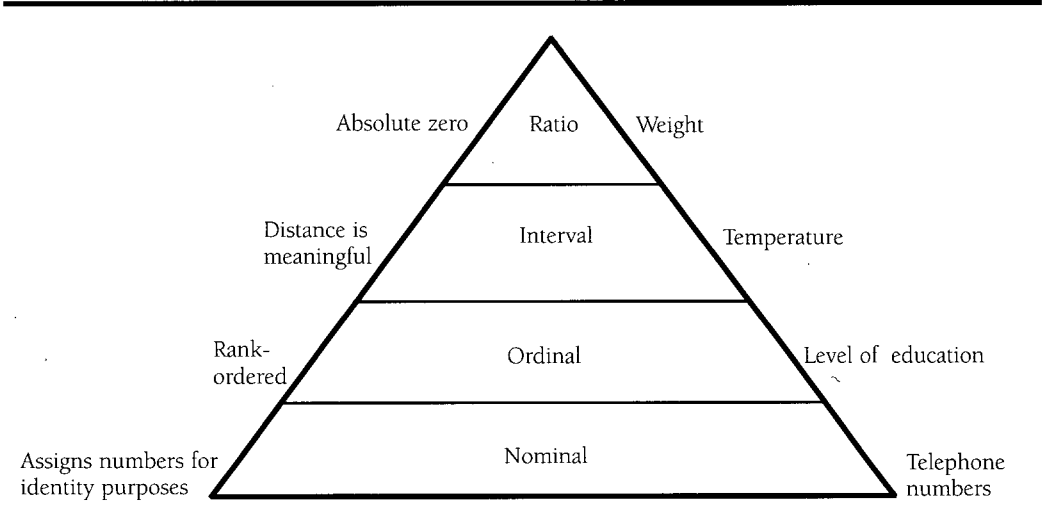
While the hypothesis represents the constructs a scientist wishes to study, the researcher must collect data to test the hypothesis. For example, a scientist may hypothesize: "Migraine sufferers who use the newer antimigraine medications have a better quality of life than those who use the older products such as butalbital." The researcher cannot directly observe quality of life, but instead must operationally define this construct. An operational definition assigns meaning to a construct or a variable by specifying the activities or operations necessary to measure it. For example, quality of life could be operationally defined as symptom-free days, a general quality-of-life instrument, or some other measure.

LEVELS OF MEASUREMENT

Once the constructs or variables have been operationally defined, the level of measurement can be determined. Measurement is important in any study because it guides the interpretation of the variable as well as the statistical analysis. Measurement is "the assignment of numerals to objects or events according to rules."[2] 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. As an example of nominal measurement, consider the concept of party affiliation. Typically, voters identify themselves as Republicans, Democrats, or Independents. For the purpose of analysis, a researcher might arbitrarily assign the values "1" for Republican, "2" for Democrat, and "3" for Independent. The values do not signify any numerical relationship among the three parties but are merely used as labels. Thus, the three parties could just as easily be labeled as "R," "D," and "I." Nominal measurement classifies objects into mutually exclusive categories. For example, in a psychiatric hospital, patients could

Figure 1. Levels of Measurement



be classified as schizophrenic, manic-depressive, or psychoneurotic. Social security numbers are another example of nominal measurement.

Ordinal measurement requires that objects in a set be ordered by rank. Distances between attributes are not equal and have no meaning beyond indicating a more-or-less relationship between categories. That is, a given category of objects has more or less of an attribute than another category. However, ordinal measures do not have equal intervals or an absolute zero. For example, surveys often code levels of education in the following manner: 0=less than high school (H.S.); 1=some H.S.; 2=H.S. graduate; 3=some college; 4=college degree; 5=post college. A value of 5 indicates more education than a value of 1, but does not suggest that a person with a value of 5 has five times more education than a person with a value of 1. Furthermore, the distance from zero to 1 may or may not reflect the same difference in education as the distance from 2 to 3. Another example of ordinal measurement is consumer product rankings, such as those seen in *Consumer Reports*.

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

Finally, 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. Since height is a ratio variable, it could be stated that someone who is six feet tall is twice as tall as someone who is three feet tall. Other examples of ratio measurement include weight and salaries measured in dollars.

As previously mentioned, the type of measurement determines the statistical analysis that can be performed. With nominal and ordinal scales, measures cannot be added together, and the variety of types of statistical tests that can be conducted is relatively limited. Interval and ratio data can be added or subtracted, and they offer a wider range of options in statistical analysis.

Exercise 1

Critically evaluate the following hypotheses.

Hypothesis A: Managed care is bad.

Hypothesis B: The quality of pharmaceutical services is directly related to the type of pharmacy (i.e., independent, chain, or mail-order).

Hypothesis C: Pharmacists should make all drug formulary decisions.

Answers: Hypothesis A does not express a relationship between two or more variables. The only variable is managed care. In addition, the hypothesis reflects a value judgment. Hypothesis B is declarative and expresses a relationship but may not be measurable because of its vagueness. Can quality be operationally defined? Hypothesis C clearly reflects a value judgment ("should") that is not amenable to empirical testing.

Exercise 2

Identify the independent and dependent variables in the following hypotheses.

Hypothesis A: Asthma patients who are enrolled in case management programs have better clinical outcomes than patients not enrolled in case management.

Hypothesis B: Therapeutic switch programs reduce drug expenditures.

Answer: In the first hypothesis, case management is the independent variable; clinical outcomes is the dependent variable. A therapeutic switch program is the independent variable in the second hypothesis, while drug expenditures is the dependent variable.

Exercise 3

For the following hypothesis, identify two ways of operationally defining the independent and dependent variables.

Hypothesis: Patients who are more compliant with their angiotensin-converting enzyme (ACE) inhibitor have lower heart failure morbidity than patients who are less compliant.

Answer: There are a variety of options for operationally defining these variables. Compliance could be defined as the number of prescription claims found in the claims database or on a scale of 1–5 (1=very complaint, 5=very noncompliant) as reported by the patient. Morbidity could be defined as hospital rates or as quality of life as measured by the SF-12, a general quality-of-life instrument.

Exercise 4

Identify the level of measurement being used in each of the following examples:

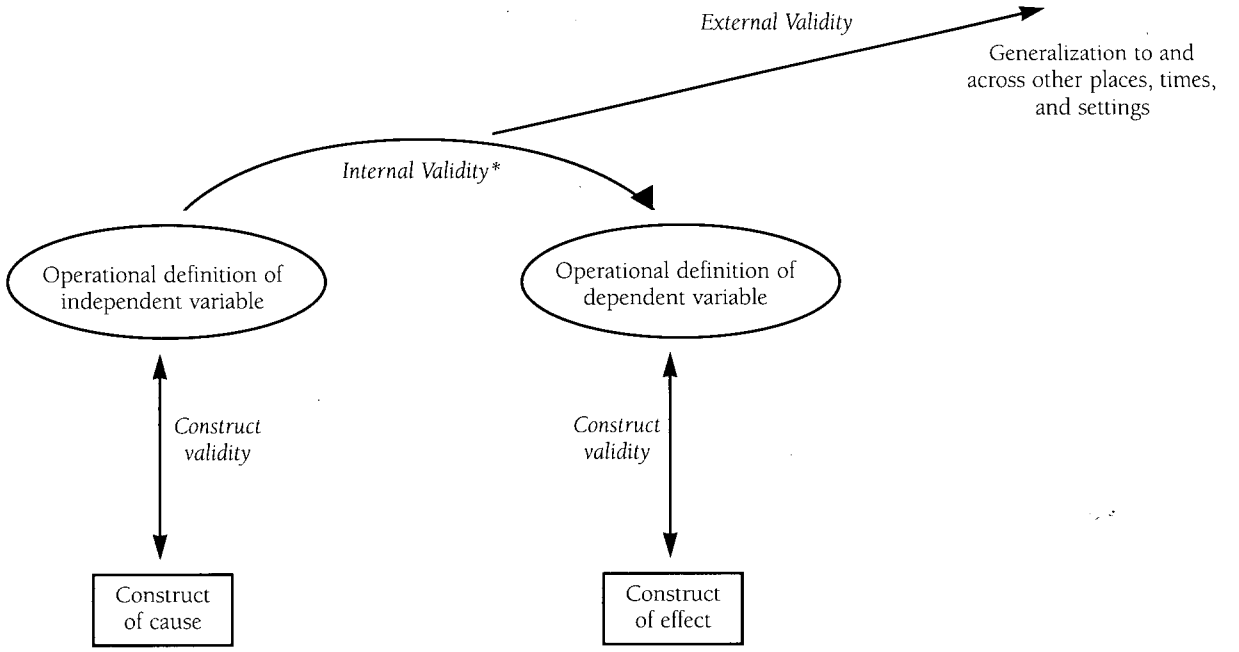
a. Race
b. Food calories
c. Telephone numbers
d. Military rank

Answers: a-nominal, b-ratio, c-nominal, d-ordinal.

VALIDITY AND RELIABILITY

Once the hypothesis has been formulated, the variables have been operationally defined, and the levels of measurement and statistical analyses have been determined, scientists must address the issues of reliability and validity. Both reliability and validity deal with the faith or trust one has in the findings and, accordingly, any conclusions that are made from the findings. Reliability is defined as the similarity of results provided by independent but comparable measures of the same object. Suppose you are measuring the quality of life of

Figure 2. Diagram of the Relationship Between Types of Validity



Statistical conclusion validity is a type of internal validity.

patients with migraine and you are using a quality-of-life instrument specific to migraine. If you measure a patient's quality of life once a week over three consecutive weeks during which the patient has no migraines, do you get the same or similar results? Inconsistency in results would suggest that the instrument is unreliable.

Another way to view reliability is to consider errors of measurement. While a certain amount of 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. When a measure is unreliable, it creates what is called random error or variance. While systematic variance tends to create errors that move in one direction, random variance tends to create some errors in the positive direction and others in the negative direction; on average, the errors balance out. A lack of reliability reduces the ability to assess relations among variables because so much error exists in the measures.

Suppose, in the previous example, the results revealed that the use of newer migraine medications was not statistically associated with improved quality of life. If the reliability of the quality-of-life instrument was poor, one could not make any definitive conclusions about the effect of medications on quality of life. Perhaps the medications do not enhance quality of life, or perhaps the quality-of-life instrument is so unreliable that true differences were undetected.

There are a number of methods for assessing reliability, each allowing for computation of a reliability coefficient. The coefficient can range from zero to 1. A higher coefficient indicates greater reliability (or less measurement error). A coefficient of 0.80 is considered by many to be the lowest acceptable standard; coefficients as high as 0.90 or 0.95 are preferred in some disciplines.

While reliability is necessary in order to draw definitive study conclusions, it is not a guarantee. Validity also must be addressed. Consider this simple example of the distinction between reliability and validity. Suppose your bathroom scale is reliable—it provides similar results over short periods of time during which your weight would not have changed. While the scale may be reliable, it is not valid if it always calculates your weight five pounds higher than it actually is.

Generally speaking, validity addresses the question: "Are we measuring what we think we are measuring?" There are four major types of validity: 1) statistical conclusion; 2) construct; 3) internal; and 4) external. This discussion will focus on the last three. Statistical conclusion validity, which is really a type of internal validity, is beyond the scope of this article.

Construct Validity

Construct validity refers to the degree to which a variable accurately reflects the phenomenon it purports to measure.[3] For example, if a researcher is studying the relationship between frequency of drug dosing and medication compliance, and compliance is measured by patient self-report, is the researcher really measuring compliance or the patient's desire to appear compliant (sometimes called social desirability bias)? Consider another example. Suppose a researcher is examining the relationship between use of inhaled steroids and acute exacerbations of asthma, as measured by emergency room (ER) visits for asthma. What is the implication if the population

being studied patronizes the ER for primary care visits rather than only acute care? Is the researcher measuring what he or she intends to measure—acute exacerbations of asthma?

Construct validity can be assessed by examining convergent and discriminant measures. Convergent measurement addresses whether the construct correlates with other concepts with which one would expect it to correlate. Discriminant validity involves examining correlations with measures from which a variable is supposed to differ. In the previous example, convergent validity can be addressed by comparing rates of ER visits to national or regional norms for comparable groups. If the observed ER visit rates were not similar to these norms, it would suggest a possible problem with construct validity.

Internal Validity

Much of science attempts to identify cause-and-effect relationships. However, simply because a correlation (association) is observed between variable X and variable Y, one cannot automatically conclude that a change in variable X causes a change in variable Y, or vice versa. Internal validity refers to the "approximate validity with which we infer that a relationship between two variables is causal."[3]

There are a number of threats to internal validity: 1) history; 2) maturation; 3) testing; 4) instrumentation; 5) statistical regression to the mean; 6) selection bias; and 7) experimental mortality. History refers to any event that occurs during the study period that can account for the findings. For example, suppose a study compares women's compliance with estrogen replacement therapy 12 months prior to and 12 months after an educational program that lasts approximately six months. Can researchers conclude that the program enhanced compliance? They could not if another major event occurred during the same time period that also could have enhanced compliance—for example, an increase in direct-to-consumer advertising about the benefits of estrogen replacement on osteoporosis.

The second threat to internal validity is maturation. It can occur when patients "mature" with the passage of time—grow older, more tired, hungrier, etc. Assume a study found that asthmatic children who were using cromolyn had improved FEV1 and reduced ER visits over a period of four years. Can we conclude that the cromolyn was responsible for the improved outcomes? Perhaps not, because it is equally plausible that many of the children grew out of their asthma and/or they learned to use their inhalers more appropriately with the passage of time.

A testing effect, another threat to internal validity, occurs when the pretest affects the post-test measurement. Suppose a teacher gave a math exam on a Monday, reviewed the difficult concepts on Tuesday, and then administered the same exam again on Wednesday. If the teacher found that exam scores rose dramatically on the second exam, would the teacher conclude that reviewing the difficult concepts had increased the students' scores? Of course not, because familiarity with the exam likely enhanced scores on the second exam. Some students remembered items and incorrect responses from the first exam.

A fourth threat to internal validity is instrumentation, which can occur if there is a change in the measuring instrument between the pretest and post-test. For example, survey administrators might become more experienced at administering a survey over time.

Statistical regression refers to a movement to the mean. It occurs because of the way people are assigned to treatment groups. For example, suppose a class of seventh-grade students is divided into two groups based on the scores the students earned on their last math exam. The group with the higher test scores attends math class as usual while the group with the lower test scores receives private tutoring. After the next exam, the math teacher finds that test scores increased, on average, for those students with private tutors and decreased for those without private tutors. Should the teacher conclude that the private tutoring program is effective at improving math performance? Even if the teacher had not provided a private tutor to the lower-scoring group, it is likely that the average test score for that group would have increased simply because some of the students in that group are better students than the first test indicated. In other words, most students have fluctuations in test scores, performing better than their true ability at times and worse at others. Accordingly, a certain percentage of the students with low scores can be expected to perform better on the next test due to a regression to the mean. The regression would work in the opposite direction among those with higher scores on the first exam. Since some of these students' scores were likely inflated by error, scores will likely fall, on average, in this group at the second test.

Selection bias is another potential threat to internal validity. It occurs when inherent differences across treatment groups are correlated with the dependent variable. For example, the researcher might find that heart failure patients taking diuretics have lower rates of hospitalization than heart failure patients taking diuretics and ACE inhibitors. It would be inappropriate to conclude that diuretics reduced the rate of hospitalization, because it is very likely that patients taking only diuretics have less severe heart failure than those also taking ACE inhibitors. Therefore, one would expect to see lower rates of hospitalization among those taking diuretics alone. When patients are not randomized to treatment groups, selection bias is more likely to be an issue.

The final threat to internal validity is that of experimental mortality, which is of concern when there are different rates of dropout across the treatment groups and dropping out is correlated with the dependent variable. As a result, the study groups comprise different types of persons by the end of the study, and these persons are different in ways that are correlated with the dependent variable. Consider a study that compares total costs of therapy for patients taking two classes of

antidepressants—tricicylic antidepressants (TCAD) and selective serotonin reuptake inhibitors (SSRIs). If the likelihood of a patient dropping out of treatment is related to the severity of the side effects and to the severity of the depression—that is, the more depressed patients are less likely to drop out of treatment, as are patients experiencing fewer side effects—a researcher might expect to see higher dropout rates for patients taking TCAD, which has higher rates of side effects than SSRIs. The dropouts would tend to be less depressed than nondropouts and, accordingly, the remaining TCAD group would be made up of more severely depressed patients than the starting group of TCAD patients. In such a situation, one could see greater overall costs in the TCAD group than the SSRI group, because the patients were simply more depressed. It would be inappropriate to conclude that the use of TCADs raised overall medical expenditures for depression.

External Validity

A third major type of validity, external validity, is the validity with which we can generalize to and across persons, settings, and times.[3] Generally speaking, a lack of external validity represents an interaction between study variables and/or subjects and the environment. Is there something unique about the study subjects and/or the setting that prevents a generalization to other subjects and settings?

For example, the use of volunteers for a pilot study of the effect of a patient education program to enhance compliance with drug therapy would constitute a threat to external validity. It is not unreasonable to believe that patients who volunteer to participate in a pilot study tend to be more compliant with their medications and, accordingly, the pilot test may suggest a greater effect from the education program than one would find in a nonvolunteer population. Another example would be telephoning subjects for a survey during the day. The characteristics of the population at home during the day do not reflect the general population and could vary systematically on the variables being studied. The key to minimizing this problem is to make cooperation as convenient as possible.

In a search of published literature, it is easy to find examples of studies whose findings cannot be generalized because of the study setting. Many studies evaluate the effect of a health management program within one health care organization, often a staff-model health maintenance organization (HMO). Given that these organizations generally provide the continuum of patient care within one organization and that many are providers of exceptionally high quality, it is questionable whether findings in this environment will hold up under less ideal conditions. The solution in this situation is to vary settings to determine whether the findings hold up under changing conditions. For example, is the program as effective in an independent practice associate model HMO?

Students often ask which type of validity is most important. There is no one right answer. Rather, the relative importance of each type of validity depends on the purpose of the study. Furthermore, there is often a tradeoff between internal and external validity. As an example, randomized controlled trials (RCTs) attempt to demonstrate a cause-and-effect relationship between drug therapy and patient outcomes. Typically, through the use of randomization and control groups (which will be discussed in the next article in this series), RCTs achieve strong internal validity. However, the external validity of these studies is usually lacking because they are conducted using the most compliant patients, the best providers, and the best centers.

Exercise 5

Read the following scenarios and describe a potential threat to validity for the study described.

> Scenario 1: A researcher identifies high-, medium-, and low-cost asthma patients. High-cost patients are assigned to an intensive educational program, while low-cost asthma patients are treated as usual. After six months, the investigator finds that the high-cost patients' expenditures have dropped since the study began, while the low-cost patients' expenditures have risen. The researcher concludes that the program is effective and should be implemented in managed care organizations (MCOs).

Evaluation: In this study, statistical regression to the mean is a threat to internal validity. If the high-cost patients had received no intervention, it is likely that the group, on average, would have incurred lower expenditures during the following six months because of statistical regression to the mean. In fact, a recent study by Lalla and Kozma[4] found that statistical regression to the mean occurs with high-cost asthma and diabetes patients. Accordingly, it is tenuous for the scientist to conclude that the education program caused the lower expenditures. Clearly, the MCO would want more convincing information before implementing such a program.

> Scenario 2: A researcher wants to determine the costs of care for patients with depression. He selects all patients in the MCO who have taken either a TCAD or an SSRI in the previous six months. He then totals their medical costs for any encounters with a diagnosis of depression over a 12-month period and finds that the costs of treating depression are lower than he expected based on published research. The research concludes that depression should not be targeted for the development of disease management programs.

Evaluation: There are two major concerns with this research, both relating to construct validity. The first is that the study included all patients taking TCAD, claiming that these were depressed patients. Because TCADs are indicated for conditions other than depression, many of the patients taking TCADs probably were not depressed. Further, the researcher included only those costs associated with a documented diagnosis of depression. It is widely known that the diagnosis of depression is frequently underreported. Accordingly, including only

those claims with a diagnosis of depression will likely underestimate the true cost of depression.

> Scenario 3: A randomized controlled trial finds that compliance with one of the FDA-approved treatments for *H. pylori* (omeprazole, clarithromycin, amoxicillin) is greater than 95%. Based on these findings, representatives of the pharmaceutical companies that manufacture these products begin convincing physicians and pharmacy directors that patients should be given this combination because it is efficacious and compliance is good.

Evaluation: Assuming the RCT was conducted appropriately, one would not dispute the finding that the drug was efficacious or even that patients were compliant. However, external validity is of concern. Patients who participate in controlled trials typically tend to be more compliant, and the trial protocol itself enhances compliance due to the rigor with which patients are monitored and encouraged to comply. This is an example of an interaction between selection and treatment (i.e., selection of more compliant patients) and between experimental arrangement and treatment (i.e., protocol encourages compliance). Accordingly, it is difficult to generalize the finding of high compliance to the population at large. It is reasonable to expect that compliance in the general population would be lower than in the RCT, especially considering that the regimen involves a large number of pills with multiple dosage frequencies.

CONCLUSION

This article has presented some of the basic concepts in hypothesis development, measurement, reliability, and validity. The next article will discuss the numerous ways to design research studies in order to minimize threats to internal and external validity.

Pharmacists are increasingly inundated with research studies on a variety of topics. While little can be done to reduce the volume of information, pharmacists can make better use of the available research by learning how to critically evaluate the studies. It takes time and practice to become proficient at critical evaluation. Although most pharmacists find it very challenging, the effort is worthwhile.

▲ References

1. Kerlinger FN. Foundation of behavioral research, 3rd edition. New York: Holt, Rinehart and Winston, Inc.; 1986.
2. Stevens S. Measurement, statistics, and the schemapiric view. Science. 1968; 161: 849-856.
3. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Boston: Houghton Mifflin Co.; 1979.
4. Lalla DV, Kozma CM. Regression toward the mean: implications for evaluations of health care interventions. Presented at the APhA Annual Meeting and Exposition, March 1998.

CE EXAM

Research Methodology: Hypotheses, Measurement, Reliability and Validity

Upon completion of this program, the successful participant shall be able to:

1. Determine and evaluate a study's hypothesis.
2. Identify independent and dependent variables in a study as well as the level of measurement of the variables.
3. Describe the concept of reliability and its importance to research.
4. Distinguish among the types of validity and identify threats to validity in a research study.

SELF-ASSESSMENT QUESTIONS

1. The level of measurement of jersey numbers is:
   a. nominal.
   b. ordinal.
   c. interval.
   d. ratio.

2. A scale with equal distances between points but without an absolute zero would be:
   a. nominal.
   b. ordinal.
   c. interval.
   d. ratio.

3. Which of the following hypotheses does not meet the requirements of a hypothesis?
   a. Students who study in groups perform better on pharmacy exams than students who do not study in groups.
   b. Group studying is the best way to prepare for pharmacy exams.
   c. Students who perform well on pharmacy exams have better job performance than students who perform poorly.

4. Patients who participate in a compliance program continue with drug therapy longer than patients who do not participate in a compliance program. In the previous hypothesis, what are the independent and dependent variables?
   a. Independent–patients; dependent–participation in a compliance program.
   b. Independent–patients; dependent–length of drug therapy.
   c. Independent–participation in a compliance program; dependent–patients.
   d. Independent–participation in a compliance program; dependent–length of drug therapy.

5. A lack of reliability:
   a. ensures validity.
   b. is associated with reduced measurement error.
   c. reflects a systematic error variance.
   d. reduces the ability to make inferences from a study.

6. Which of the following is <u>not</u> a type of validity?
   a. Construct
   b. Internal
   c. Nominal
   d. External

7. The ability to make generalizations to other settings is an issue of:
   a. statistical conclusion validity.
   b. construct validity.
   c. internal validity.
   d. external validity.

8. Developing operational definitions that reflect the variable of interest is an issue of:
   a. statistical conclusion validity.
   b. internal validity.
   c. construct validity.
   d. external validity.

9. When two treatment groups are inherently different on a dimension that is expected to correlate with the dependent variable, the threat to validity is termed:
   a. statistical regression to the mean.
   b. selection bias.
   c. instrumentation.
   d. maturation.

10. A quality-of-life instrument that provides consistent measures over time but always overestimates patients' quality of life is said to be:
   a. reliable but not valid.
   b. valid but not reliable.
   c. neither reliable nor valid.
   d. both reliable and valid.

See text of article beginning on page 382 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 233-000-98-004-H04 in AMCP's educational offerings.

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.

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 233-000-98-004-H04. This offer of continuing education credits expires August 31, 1999.

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

	A	B	C	D
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

Date _____

Social Security # _____
For Identification Purposes Only

Name _____ Work Phone # _____
Last First Middle

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 _____