Sources of Bias in Research Design

Editors' note: This article is the seventh in a multipart series designed to improve the knowledge base of readers, particularly novices, in the area of clinical research. A better understanding of these principles should help in reading and understanding the application of published studies. It should also help those involved in beginning their own research projects.

The previous installment in the Basics of Research Series addressed sources of bias related to the measurement of research variables. In contrast, this issue addresses sources of bias within the research design, focusing specifically on internal and external validity. Note that although the word "validity" is common to both, validity here refers to validity of the entire study, rather than an individual variable.

Threats to Internal Validity

Many factors potentially can introduce bias into a research study. *Internal validity* is the degree to which changes or differences in the dependent variable (the study outcome of interest) can be attributed to the independent variable (intervention or group differences). In other words, are the results really true within the sample examined? This question is of great importance because it can determine the degree to which the results should guide future practice, as well as the likelihood that the results of the study will be considered publishable. Internal validity is of greatest concern when a study is attempting to determine causality.[1] Does the independent variable "cause" the dependent variable?

Extraneous (sometimes confounding) variables are one common source of bias that can influence study outcomes, even if not a primary focus of the study itself. Extraneous variables come in many different forms and have various sources. Extraneous variables can be an unrecognized cause of the study results, even if the variables are not measured or even identified in the study. For example, a researcher may conclude that the intervention under investigation was the cause of the outcome, when in actuality, the cause was some other extraneous variable. Several potentially confounding extraneous influences are discussed.

History (ie, historical changes in the environment) is one important type of extraneous variable. Bias due to *history* occurs when changes in the outcome variable, over time, are attributed mistakenly to the study intervention rather than to other changes in the environment. For example, a flight program may institute a new educational program and attribute an improvement in quality of patient care to this intervention. However, the researchers must be concerned that some other change in the environment was the actual cause of the improved quality of care. The addition of a second flight nurse to the team, a modification in onboard equipment, or some other change could be playing a part in the noted improvement in quality. As with many of the identified threats to internal validity, the use of a control group may help to minimize this possibility. If both groups experience the same history and only the experimental group demonstrates improvement in the quality of care, the researchers can have more confidence that their intervention was the cause of the change in quality.

Another category of extraneous variable is *maturation*, which refers to changes in the dependent variable as a result of normal, intrinsic changes over time. For example, after a surgical procedure, pain naturally decreases over time. Thus, an investigator should not necessarily attribute a decrease in pain after surgery to an intervention because of the potentially confounding effect of maturation. Repeated measurement of the dependent variable (outcome variable) can help to control for the effects of history or maturation. Analysis of trends over time can help identify changes attributable to the intervention versus changes that would occur even without intervention.

Instrumentation is another category of extraneous variables that can be a threat to internal validity. Instrumentation issues are related to the quality of the instrument or how it is used in the study. Instrumentation can be caused by a change in how an instrument functions between time 1 and time 2 (blood pressure cuff no longer calibrated accurately) or by a change in how the researcher uses the instrument because of their own increased skill level. With instrumentation problems, a given instrument could give different results, even when the variable itself has not changed.

A related threat to internal validity is *testing*. Bias due to *testing* may result when performance improves because of practice on the "test" (you get better the second time you try

most activities) or remembering the right answers, rather than the effect of the independent variable. An example of bias resulting from *testing* could be testing flight crew performance (speed, error rates) in a video game simulation, before and after a night shift. Changes in performance might be attributable to learning how to play the game rather than the effects of having worked all night. Bias due to *testing* might also be caused by the direct effect of the instrument itself. For example, the contemplation needed to respond to a question related to ethical/unethical care situations may alter the subject's viewpoint related to those issues.

Loss of subjects (*mortality*) during the study can bias outcome results. If some variable besides chance affects loss of subjects in one group more than in another, it may be this variable and not the intervention causing the observed outcome. For example, if the subjects who do not like your approach or who do not respond successfully to your treatment preferentially drop out, the final study sample could indicate an erroneous benefit from your intervention.

Finally, the *method of assignment* of subjects to experimental and control groups could influence the outcome of the study. If, for example, all subjects were assigned to the experimental group until that group had enough subjects, and the second half of the volunteers were assigned to the control group, the first group might be significantly different from the second group even before the intervention was applied. Individuals who volunteer early may be inherently different than those who delay and volunteer later or after much encouragement. Consequently, the final study results may reflect more beginning group differences rather than true effects of the intervention.

The most successful method for dealing with assignment difficulties is to randomly assign individuals to the groups. *Random assignment* minimizes preexisting bias in the subject assignment, but it may not always be effective. For example, even if the investigator flips a coin to assign subjects to groups, the investigator could be unlucky and get 14 out of 20 heads rather than an even distribution of heads and tails, or could by chance get more men in group 1 than in group 2.

If the investigator needs to ensure that subjects with a particular characteristic are distributed evenly, subjects can be stratified, based on a given characteristic, before assignment to groups. This approach is not always necessary but can be important in some studies. For example, if gender is expected to make a difference, the investigator can ensure that an equal number of men and women are assigned to each group in the study protocol design.

If the investigator does not wish to stratify on a potentially confounding variable, a more homogeneous sample may be used. For example, if gender is expected to cause differences in the outcome, the investigator could study only men or only women by including that in the study inclusion/exclusion criteria.

Control or adjustment for extraneous variables is an important consideration when designing a study. Many of these

> "The most successful method for dealing with assignment difficulties is to randomly assign individuals to the groups."

issues can also be addressed with statistical techniques at the conclusion of the study. However, when possible, it is best to try to minimize their effects in the design phase. For example, the investigator can try to control for the effects of testing by adding additional groups that do not receive a pretest (eg, Solomon 4 group designs); the effects of history and maturation can be controlled for by the use of control groups who are similar to the experimental group except in regards to the intervention.

Threats to External Validity

In contrast to internal validity, external validity is the degree to which the results can be applied to others outside the sample used for the study. In many studies, the results can be generalized only to the type of individuals specifically included in the study or even that specific sample, because of something unique about that group or situation. A number of external variables, such as the study environment or conditions, or even the investigator's presence, may influence the study subjects or the measurements. For example, if a study of an educational training session was done in a hot room at the end of the day, the results may be applicable only to tired, irritable subjects but not to subjects who do not have these characteristics.

The *Hawthorne effect* is one factor that may influence external validity. The Hawthorne effect occurs when subjects respond in a different manner just because they are being observed. For example, it may be the influence of having a researcher paying attention to the transport program that causes the subjects to change their attitude and performance rather than as the result of the study intervention (the independent variable). Once the researcher is removed from the situation, performance and attitudes may return to pre-study levels.

A related issue is the *novelty effect*. Subjects may alter their performance, become more engaged, or change attitudes just because something is new. Once the novelty wears off, their behavior may return to pre-study levels. In this case, the results would only apply to subjects who were new to the intervention.

Repeated measurement of anything can have an effect on subjects. If study subjects are exposed to a large number of questionnaires, observations, etc., they may become tired of the procedures or are so accustomed to them that their performance is altered. Consequently, data obtained from such a complex study may apply only to others involved in similar complex conditions.

A final threat to external validity would result in the case of an *interaction between history and treatment effect*. In this case, there is something unique about the time and place that makes the treatment more (or less) effective, and thus not widely generalizable. For example, an intervention to increase safety behaviors may only be effective for teams who have recently experienced an accident; the intervention may be ineffective with teams not having had this experience recently.

Conclusions

The goals of most research are to investigate a small sample of subjects, determine the outcome of the study intervention, and be able to generalize the findings to a broad group (eg, the general population). Internal validity is the degree to which the findings accurately reflect reality. External validity is the degree to which the results can be generalized to other groups. The ability to limit sample size while maintaining internal and external validity will decrease the time and cost of most studies.

However, balancing these needs is often a challenge. Changes to increase one will often decrease the other. Samples recruited using nonrandom techniques are easier to obtain, but limit the external validity of the study because of greater potential for bias in subject selection. Restrictive inclusion/exclusion criteria can be used to maximize the internal validity of the study by minimizing potential confounding variables. But if they excessively narrow the population of interest, they can decrease the external validity of the results. Thus researchers must maintain an awareness of the need to balance all considerations related to internal and external validity when designing a study.

No one design will meet all needs, at all times. However, a researcher with a strong understanding of the issues will be able to design an effective study and defend the resulting conclusions.

Reference

1. Burns N, Grove SK. Practice of nursing research: conduct, critique, & utilization. St. Louis, MO: Elsevier Health Sciences; 2005.

Cheryl Bagley Thompson, RN, PhD, is an assistant dean of informatics and learning technologies and director of University of Nebraska Medical Center informatics, College of Nursing, in Omaha, Nebraska. She can be reached at cbthompson@unmc.edu. *Edward A. Panacek, MD, MPH, is professor of emergency medicine and clinical toxicology at the UC Davis Medical Center in Sacramento, California.*