Research Study Designs: Experimental and Quasi-Experimental

> *Editors' note: This article is the third 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 first article in this series discussed developing an area of general interest and generating a proposed research question or hypothesis. The second article discussed reviewing the relevant body of literature on the subject and confirming that the research question is an appropriate one. The next step is planning the proposed research project. Translating a research idea into an actual project requires an understanding of research study designs. This article discusses the experimental research designs that are most appropriate for clinical investigations; the next article will address nonexperimental designs. Having an understanding of the full spectrum of research designs helps the investigator to select the design that would be most appropriate to answer an individual research question.

The study design is the general plan for setting up and testing a specific hypothesis or research question. Overall, the design directs the who, what, when, and how of the research project. Think of the design as the basic foundation or infrastructure for the project. Built on top of the design are the specific elements of the study protocol itself, which will be discussed in detail in a future article. A common misconception is that for every research project there is one single best design to answer that research question. In reality, there are usually multiple research designs that can be used to approach a given research question. However, each one of the designs has unique advantages and disadvantages.

Decisions regarding selection of a research design for a given project represent a compromise between the goal of rigorous scientific integrity versus limited resources and clinical reality. A common assumption is that we should always strive to achieve the "gold standard" of a "prospective, randomized, blinded, controlled clinical trial." However, such trials are the most expensive type of study to perform and often are not the most appropriate to answer certain types of research questions. Sometimes they are not even an option for ethical or legal reasons.

Another common misunderstanding is the belief that retrospective studies are "worthless" from a scientific standpoint. Although they do have real limitations, many research ques-tions can be answered only through retrospective study designs. Every different research design has a potential application in a given setting, and each design also has limitations. The large prospective, rigorous, clinical trials, which we now take for granted, are a relatively recent development. The earliest multicentered, randomized, controlled clinical trials found with PubMed were in 1966.[1,2] During the past 40 years, the important elements of research study design and clinical trial development have been refined further and are now well established.

Many different research study designs exist. To best understand how the various designs interrelate, a classification system is useful. Multiple classifications are available, and individual textbooks use different systems, sometimes resulting in confusion. Examples include classification by "time frame of data collection" (ie, prospective, retrospective, ambispective), "assignment of study groups" (ie, randomized vs. nonrandomized), and "degree of masking" (ie, blinding vs. not).

One of the simplest systems simply divides the research designs into retrospective or prospective categories. However, confusion often exists in relationship to these terms. Retrospective studies are those in which the events of interest all occurred before the onset of the study. Even when the research question and design are generated prospectively, before starting the study, if the events that the investigation will be studying have already occurred, then the study is ret-rospective. In prospective studies, the events of interest have not yet occurred when the study begins. This provides the ideal opportunity to maximize the accurate collection of relevant data.

The research design classification system that is in most common use is "classification by scientific rigor." This system categorizes the research designs based on levels of overall sci-entific integrity. In other words, it takes the perspective of: "If this research design is used for this study, how scientifically valid are the results?" True experimental designs are those that have a structure that generally yields the most valid results.

Quasi-experimental designs are one step down and have a moderate level of scientific validity. Nonexperimental designs are those research designs that, by virtue of their overall structure, give results that generally do not have strong scientific validity. The ability to draw firm conclusions from the study results is directly proportional to the level of scientific validity of the design. This classification system will be used throughout the remainder of this article.

Study Variables

First, some basic definitions should be reviewed (Table 1). Three important variables apply to all designs. The *independent variable* is the specific study "intervention" or prediction variable. For example, treatment A versus treatment B or diagnostic technique A versus diagnostic technique B are two common examples of independent variables in health care research. It need not be an active intervention; it could be an already existing exposure or trait that could serve as a predictor variable. An example would be the presence of pre-hospital hypotension versus not, in blunt trauma patients. That variable might predict their outcome or injury pattern (or not). The *dependent variable* is the main study "outcome" that is being measured. In general, the dependent variable is believed to be affected by the independent variable (eg, mortality rates, injury patterns, or complication rates). Either the independent variable or the dependent variable could be the parameter of greatest interest in a given study, depending on the research question being asked. The most common study model is to manipulate the independent variable (eg, to give drug A or drug B or select study patients based on the presence or absence of trait) and then measure their associations with subsequent outcomes (the dependent variable).

Extraneous variables are other factors that could exert outside influences. They are not being introduced or controlled directly as part of the research design but might have an important impact on the results. Such extraneous variables may distort the relationship between the independent and dependent variables and may be unequally distributed among the study groups. For example, if the study involves treatment of asthma, then smoking history might be an important

potential extraneous variable. If a larger number of smokers are found in the treatment group, then the effect of the treatment may be blunted because of preexisting compromised lung function from the smoking. Extraneous variables should be considered carefully because they can significantly alter the study results and sometimes even invalidate the entire study. Predicting all of the important extraneous variables in advance is difficult. Because of this, recognizing that some research designs are inherently better at controlling for extraneous variables than others is important.

Study Design Elements

Three other elements are important in understanding study designs: manipulation, control, and randomization. *Manipulation* is the ability of the researcher, and thus the design, to interact with the study subjects to effectively direct the independent variable. Prospective interventional studies clearly demonstrate manipulation, as the researcher manipulates the intervention such that one group receives it (eg, the new therapy) and the other does not. In contrast, researchers conducting purely observational studies do not interact with the study subjects in this way and simply record data as they occur, without any "manipulation" of the variables.

Control refers to whether the researcher has influence over the study environment itself and thus the ability to limit potentially confounding variables. For example, does the study design direct or influence patient care measures that are not the independent or dependent variables? This could include when and how to make study measurements. More importantly, for interventional studies, does the study protocol control other care being delivered? For example, in a study of high dose versus standard dose albuterol for severe asthmatics, is the administration of O_2, steroids, antibiotics, and so forth, also standardized by the protocol? Prospective studies generally have some degree of control. However, retrospective studies never do, unless they were part of a prospective protocol in use at the time.

The third element in research design is randomization. *Randomization* refers to how subjects are assigned to study groups. Designs in which randomization is used provide each subject with a known probability of being assigned to each of the study groups (eg, experimental or control). In most studies the probability of assignment to the two groups is equal. However, if a larger experimental group is needed for a specific reason, the randomization process might, for example, assign two subjects to the experimental group for every one subject assigned to the control group.

Several methods can be used for ensuring random assignment to groups. Drawing numbers out of a hat (1 = control, 2 = experimental), use of a random number table or computer program, or flipping a coin are common methods for ensuring random assignment. Random assignment is not the same as random selection. Random selection is the process of sampling or choosing subjects to enroll in the study. It is part of the study protocol but is not a component of the general study designs, per se. Subject selection is important to the validity of the study but will be discussed in a subsequent article addressing sampling techniques. Only random assign-

Table 2. Classification of Research Designs by Degree of Scientific Rigor

True experimental
- Have all three design elements
- Are always prospective
- Have high scientific validity

Quasi-experimental
- Only have one or two design elements
- Have manipulation or control
- Generally lack randomization
- Are generally prospective in nature
- Are moderate in scientific validity

Nonexperimental
- Have one or none of the core design elements
- Lack manipulation and randomization
- May also lack control
- Are generally retrospective
- Have the lowest scientific validity

ment impacts on the classification of the study as experimental, quasi-experimental, or nonexperimental.

Table 2 summarizes study designs that use the classification system "degree of scientific rigor." True experimental designs have all three of the primary characteristics (ie, manipulation, randomization, and control). They can be thought of as being most similar to a highly controlled laboratory experiment. As such, these designs have the most safeguards against sources of bias and therefore the greatest degree of overall scientific validity. Quasi-experimental designs are missing one or two of these elements. They have the element of manipulation or control but rarely randomization. These designs have the next highest level of scientific validity after "true experimental" but have scientific limitations. As a result, their results are more prone to bias and therefore have less validity.

Nonexperimental designs always lack randomization and usually one or both of the other primary characteristics as well. These studies are generally "ex post facto," that is, retrospective-type designs. Because the study elements of interest have already occurred, it is not possible to randomly assign subjects for the purposes of this specific study. This is true even when using a prior existing data set from a prospective randomized study that was originally conducted for a different reason. As a result, they have the lowest level of scientific validity and, to variable degrees, the findings are always open to question. It is important not to overstate the conclusions from studies using a research design from this category.

True Experimental Research Designs

True experimental research designs are always prospective in nature. A true experiment can effectively argue a proven cause-and-effect relationship. They are the most effective at demonstrating efficacy of a new intervention or treatment. To bring a new pharmaceutical product to market, the Food and

Drug Administration (FDA) will require compelling evidence of efficacy shown in a true experimental research design study (ie, a prospective, randomized, controlled, blinded clinical trial). On the downside, these types of studies are the most demanding in terms of time, cost, and other resources.

By their nature, true experimental designs tend to be tightly focused, and each study generally can look only at a narrow, highly specific research question. As such, they are not appropriate when a field of investigation is immature and the research questions still broad in nature. They are not appropriate designs for broad questions. They should be used to answer focused questions supported by prior work. Otherwise, a great deal of resources can be spent barking up the wrong tree. Before using a true experimental design, preliminary work should already have been performed, using less rigorous designs that support asking the focused research question.

Many true experimental research designs exist, but they have generally similar structures. The most common ones are outlined in Figure 1. In those descriptions, observation (O) can be any measurement or other data collection activity. Intervention (X) is the manipulation of the study intervention such as drugs, new diagnostic studies, and so forth. The control or comparison intervention (C) could be placebo or current standard care, depending on the study. The classic *two-arm design* is that which would be used to study an intervention, such as a new drug. The design compares the new therapy with a placebo or other control and examines the outcome of patients randomized to the two groups, one that receives the new intervention and one that receives the control comparison.

A *three-arm study* is very similar but compares two different drugs with control. The extended follow-up design takes the principles of the two-arm design and makes multiple subsequent observations over a longer period (eg, hospital admission rates, length of stay, relapse rates). This design takes

Figure 1. True experimental designs.*

Two Arm
R O X O
R O C O
Three Arm
R O X^1 O
R O X^2 O
R O C O
Follow-up
R O X O^1 O^2
R O C O^1 O^2
Factorial
R O X^1 O
R O X^2 O
R O X^1 X^2 O
R O X^2 X^1 O
Crossover
R O X^1 O X^2 O
R O X^2 O X^1 O

R = randomization, O = observation, X = active intervention(s), C = control intervention

Table 3. Quasi-experimental Designs

Group sequential study
- Takes one group and exposes it to multiple interventions, in a set order. Controls well for extraneous variables. Useful when the number of study subjects is limited.

Cohort study (prospective or ambispective)
- Selects a study group with common characteristics and a control group that is similar but without the variable of interest. Used when the independent variable is uncommon or randomization is not possible. Follows the groups for outcome.

longer to perform but can provide clinically important outcome data. The *factorial design* looks at the effect of multiple interventions, both individually and in various combinations. For example, two different drugs can be evaluated for individual effects, as well as their cumulative effects, depending on the order in which they are given. This design is attractive to examine each intervention, as well as combined sequences of therapies. However, it can be very resource expensive because the additional study arms require more total study patients.

The *crossover design* has the advantage that each subject becomes their own control, thereby directly controlling for most extraneous variables. The disadvantage is that it requires a lengthy period of study and therefore is generally not appropriate for the acute care setting. Crossover designs are used more commonly in a clinic setting with long follow-ups and require an adequate "wash-out" period between successive interventions. Otherwise, unintended interactions may occur between each intervention (ie, between the independent variables).

Separate from their cost, true experimental designs have other disadvantages. Random assignment of patients to the respective study groups may not always be ethical or even possible. For example, if a research goal is to study the medical effects of cocaine use, it obviously would be impossible to assign patients randomly to cocaine use. Such a study cannot be done in humans using a "true experimental" design. True experimental designs also are impractical for clinical events or conditions that are very uncommon or in settings in which the clinical environment cannot be controlled. As we will discuss, other study designs are more appropriate for those circumstances.

Quasi-experimental Designs

Quasi-experimental designs lack one or two of the study elements. They often have manipulation of the independent variable or control of the study setting, but rarely have randomization. Although the degree of scientific validity is not as high as in true experimental designs, for some research questions these are the best and most valid designs available. Quasi-experimental designs can help to validate treatment methods or establish potential associations. However, because they usually lack random patient assignment to study groups, there is an increased potential for bias, or confounding, and study validity is compromised. As such, these designs can sometimes be used as a stepping stone to establish the rationale for subsequent, focused, true experimental designs in the same field.

Quasi-experimental designs are generally less expensive than true experimental designs and are sometimes the best or only realistic option for ethical or other reasons. The most common quasi-experimental designs are listed and outlined in Table 3. The *group sequential design* is sometimes also called a "single group time series." A single population of subjects is selected and used as its own controls as it goes through a series of observations and interventions, all in the same order. The advantages of the design are two: First, the design controls for potential extraneous variables by using each patient as his or her own control, much as the crossover design did. Second, the design requires fewer subjects and therefore has an application in settings in which the number of potential study candidates is limited. The trade-off is that scientific validity is lower because randomization is absent. All subjects undergo the interventions (experience the independent variable) in the same order, so blinding is not possible. In addition, particularly in the acute-care setting, it can be very difficult to track study subjects for lengthy periods and put them through a series of sequential interventions. It is more relevant to scheduled laboratory-type experiments than to actual patient clinical research.

Cohort studies are among the most popular of the quasi-experimental type. These studies are sometimes called follow-up or longitudinal studies. The term *cohort* comes from the old Roman armies, where it was used to describe a large circumscribed group of similar or identical soldiers (eg, all foot soldiers, cavalry, or archers). They then moved together, as a group, through space and time. However, the term *cohort* is frequently misunderstood in clinical research, in part because there are multiple different variations of cohort studies, and it can be confusing. Cohort studies may be single group or multiple group.

Analytic and observational studies. Even though classic cohort studies are prospective, retrospective cohort study designs exist.

In an analytic cohort study, two similar populations are selected. One group has the independent variable of interest at the time of study entry and the other group does not. The groups can be selected concurrently or sequentially, but in either case, the patients are followed and observed for development of the dependent variable (outcome) of interest. For example, an analytic cohort study would be ideal for examining the effects of cocaine use. The study group are all proven cocaine users. The control group should consist of closely matched subjects who are clearly not using cocaine. Each group would then be followed and data would be collected over time. This is why cohort studies are sometimes called fol-

low-up or longitudinal studies. Sometimes the term "two group cohort study" is used to describe this design.

Single group cohort studies (eg, the Framingham study) enroll a group of subjects (cohort) with some common factor and follow them over time. The common factor could be year of birth (a birth cohort), living in the same city, or workers from the same factory (presumably with some industrial exposures). This is ideal for measuring incidence rates or prevalence and looking at potential associations (eg, cholesterol level and CAD). Because subjects are selected based on their exposure (the independent variable), cohort studies are ideal for studying the effects of relatively rare exposures on outcomes and measuring incidence rates. Sometimes they are also the best option for ethical or other reasons. Cohort studies are generally prospective but can be performed retrospectively if the relevant data are available. When they use a prospective study design, the data collection is generally much more complete and, therefore, the scientific validity is higher than for retrospective studies. However, in addition to some limitations in scientific validity, cohort studies also have other disadvantages. Prolonged follow-up periods, when necessary, can be relatively costly and difficult to perform. When the outcomes of interest are rare, there are better study designs (eg, case control) to use. Overall, results of cohort studies also are better for establishing associations between variables than proving true causality, because they lack randomization, which increases the risk of bias.

Summary

A core set of clinical research designs have stood the test of time and are repeatedly used in most studies. No single research design is best to answer all research questions, and every research design has appropriate applications. The next article will discuss nonexperimental designs.

References

1. Grant RHE, Keelan P, Kernohan RJ, Leonard JC, Nagekievill L, Sinclair K. Multicenter trial of propranolol in angina pectoris. Am J Cardiol 1966;18:361-365.
2. VA Cooperative Study of Atherosclerosis, Neurology Section. An evaluation of estrogenic substances in the treatment of cerebral vascular disease. Circulation 1966;33:113-119.

Other Suggested Readings

Cook TD, Campbell DT, Shadish WR, Shadish W. Experimental and quasi experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2001.

Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. St. Louis, MO: Mosby–Year Book Inc; 1996.

Hulley SB, Cummins SR, Browner WS, et al. Designing clinical research. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2001.

Stiell IG, McKnight RD, Greenlier GH, McDowell I, Nair RC, Wells GA, et al. Implementation of the Ottawa ankle rules. JAMA 1994;271:827-831.

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