# Experimental Research Design

OJMARRH MITCHELL

The design of research is fraught with complicated and crucial decisions. Researchers must decide which research questions to address, which theoretical perspective will guide the research, how to measure key constructs reliably and accurately, who or what to sample and observe, how many people/places/things need to be sampled in order to achieve adequate statistical power, and which data analytic techniques will be employed. These issues are germane to research of all types (exploratory, explanatory, descriptive, evaluation research). However, the term "research design" typically does not refer to the issues discussed above.

The term "experimental research design" is centrally concerned with constructing research that is high in causal (or internal) validity. Causal validity concerns the accuracy of statements regarding cause and effect relationships. For example, does variable 1 cause variation in variable 2? Or does variable 2 cause variation in variable 1? Or does variable 3 cause variation in both variables 1 and 2? And what is the magnitude of the causal relationships among the variables? Thus, research design as used herein is a concern of explanatory and evaluation research but generally does not apply to exploratory or descriptive research.

The importance of making causal inferences in criminology is hard to overstate. A central issue in many criminological debates concerns whether correlates of offending are causally related to offending. The correlates of offending are well known: bad parenting, deviant friends, prior delinquency behavior, youthful age (i.e., adolescents and young adults), being male, deviant attitudes, personality traits such as impulsivity and psychopathy, and so forth. Criminologists largely agree on these correlates of offending. Yet, "correlation does not imply causation." The field of criminology is filled with debates about which of these relationships are causal in nature. Perhaps the best known of these debates focuses on association between deviant peers and offending. Social learning theorists assert that having numerous, close relationships with those involved in deviance *causes* one's own level of deviance to increase. On the other hand, social control theorists argue that this relationship is noncausal; instead, the positive relationship between having deviant peers is the result of "homophily" (the tendency of individuals to associate with similar others) – "birds of a feather flock together." Likewise, there is disagreement over whether the relationship between prior offending and future offending is causal. Theorists such as Gottfredson and Hirschi (1990) argue that this relationship is spurious, as both prior and future offending are caused by low self-control. Other theories, such as Sampson and Laub's (1993) age-graded theory of informal social control, assert that involvement in crime and contact with the criminal justice system increase the likelihood of future offending because these experiences diminish bonding to important sources of informal social control (e.g., marriage, employment).

Debates concerning causal inference are not confined to theory. The effectiveness, or causal effect, of many criminal justice-based interventions on measures of offending are hotly debated. Evaluations of criminal justice interventions (e.g., reentry programs, drug court, and domestic violence programs) often find that program participants have less recidivism than nonparticipants. Yet, most evaluations have difficulty proving that the observed differences were actually *caused* by program participation.

Simply put, research design is a central concern in criminology because carefully designed research that is implemented with high fidelity can establish causal validity/causal inferences.

## Criteria for Establishing Causal Inferences

The three classic criteria necessary to support a causal inference, according to the philosopher

*The Encyclopedia of Crime and Punishment*, First Edition. Edited by Wesley G. Jennings.

DOI: 10.1002/9781118519639.wbecpx113

John Stuart Mill, are: (1) association (correlation), (2) temporal order, and (3) nonspuriousness. The criterion of association requires that there is a systematic relationship between the cause and effect variables. This criterion is by far the easiest to determine. The second criterion of temporal order is a bit more complicated. The temporal order criterion requires that the cause, or more precisely variation in the cause variable, must occur before the observed variation in the effect variable. The third criterion of nonspuriousness is by far the most difficult to achieve. This criterion requires that the observed relationship between the cause and effect variables must not be due to other omitted or unmeasured third variables. Using the relationship between delinquent peers and offending as an example, this criterion requires that this relationship cannot be due to homophily or any other potential explanation. Because there are usually many, many potentially relevant third variables and many of these third variables are unobserved, the criterion of nonspuriousness can be quite difficult to achieve.

## Types of Experiments

Shadish, Cook, and Campbell (2002) define an *experiment* as "a study in which an intervention is deliberately introduced to observe its effects" (p. 12). Shadish and colleagues distinguish two broad types of experiments: *randomized* experiments and *quasi* experiments. The central difference between these two types of techniques is the use of *random assignment* to the levels of the hypothesized cause variable.

The hallmark of all randomized experiments is the use of random assignment to experimental conditions. In randomized experiments, research subjects are randomly assigned to different levels of the hypothesized cause variable (i.e., experimental conditions) by the researchers. Random assignment can be achieved in many different ways, such as by flipping a coin, using a table of random numbers, or using numbers randomly generated by a computer. The method of randomization is largely arbitrary, but the use of some form of randomization is the crucial element of a randomized experiment.

Randomized experiments come in many forms or designs. The most common form of a randomized experiment involves randomly assigning research subjects, all of whom have been screened for eligibility, to either the treatment group that receives the experimental intervention of interest or the control group that does not; the control group, instead, typically receives no treatment, standard care, or a placebo. Randomized experiments involving the use of a no-treatment control group are often referred to as "randomized controlled trials." Randomized controlled trials are considered by many to be the gold standard of evaluation research for their high causal validity. There are many variations of this basic design. One common variation involves multiple treatment groups that receive varying doses of the experimental intervention. Another common variation involves "blinding" – procedures designed to prevent research subjects, treatment providers, and/or researchers from knowing which experimental condition a research subject was assigned. Double-blind randomized control trials typically attempt to prevent research subjects and researchers from learning which research subjects were assigned to the control group, until after all data have been collected. Blinding is intended to prevent various kinds of bias from contaminating the research results. Randomized experiments are increasingly common in criminology (see, e.g., Farrington & Welsh, 2005), but double-blind randomized experiments are extremely rare.

Quasi experiments do not use randomization to assign research subjects to experimental conditions; instead, some other method of assignment is utilized. Often research subjects voluntarily choose to participate or not to participate in the treatment of interest. Thus, the actions and wishes of the research subjects typically affect assignment.

Quasi experiments utilize a wide variety of designs. The two most common involve one-group and two-group designs. The simplest and least rigorous quasi-experimental research design involves one group of research subjects who participated in some treatment of interest. These research subjects are observed before and after the administration of treatment of interest. And the observed changes in the outcome of interest are causally attributed to participation in the treatment. Another widely used quasi-experimental design involves the use of two groups. Typically, two-group quasi experiments involve a comparison group that does not receive the treatment of interest and a treatment group that does receive the treatment. These groups are compared, often while controlling for any observed differences, and the remaining differences are causally attributed to the treatment.

Randomized experiments and quasi experiments are capable of clearly establishing the first two criteria for causal inferences (association and temporal order); yet, they differ sharply in their ability to establish nonspuriousness. Randomized experiments are able to convincingly establish nonspuriousness because of their use of random assignment. Random assignment ensures that research subjects will be equal in expectation on all variables – *both observed and unobserved variables* – prior to the administration of the experimental intervention. The phrase "equal in expectation" does not mean that the research subjects assigned to each of the experimental conditions will be perfectly equal on all variables. Instead, equal in expectation means that if we could repeat this assignment process an infinite number of times, the population means on all variables would be equal for each of the experimental conditions. Therefore, any differences between research subjects assigned to the various experimental conditions are due to chance. Because there are no *systematic* differences between the experimental groups on any variable besides the experimental condition, randomized experiments are able to rule out all potential third variables as alternative explanations for the observed differences on the outcome variable(s) of interest.

Quasi experiments have much greater difficulty in establishing nonspuriousness. In quasi experiments the actions and/or wishes of those involved in the research affect which experimental condition they eventually receive. This is highly problematic, as research subjects who choose to participate in a particular level of the experimental condition often differ from other research subjects on observed and/or *unobserved* variables. If research subjects in various levels of the experimental condition (e.g., program participants vs. nonparticipants) differ only on observed variables, then it would be easy to control for these observed differences by using statistical techniques such as multiple regression. However, in the absence of random assignment, how does one establish convincingly that participants and nonparticipants differ only on observed variables? It stands to reason that if the groups differ on observed variables, then they also differ on *unobserved* variables as well. Further, even if participants and nonparticipants are equal on observed variables, this does not mean that these groups are also equal on important *unobserved* variables. This is *the* crucial issue, because it is these unobserved differences that cause selection bias. *Selection bias* refers to inaccuracies in the estimated relationship between variables that is caused by omitted or unmeasured variables. Because quasi experiments do not establish that research subjects are equal in expectation on all variables, especially unobserved variables, prior to the administration of the experimental intervention, *selection bias is a persistent problem in quasi-experimental research designs*.

In the language of research methods, in randomized experiments, the assignment of research subjects to experimental conditions is *exogenous*. Exogenous in this context means outside or *external* to everyone involved in the experiment including the research subjects, treatment providers, and researchers – only randomization affects experimental assignment. Research subjects have no influence on which level of the experimental condition they will be assigned. However, in quasi experiments, the actions and wishes of those involved in the research including research subjects, their families, treatment providers, criminal justice officials, and researchers among many others may affect assignment; and therefore, assignment in quasi experiments is *endogenous* – meaning that the assignment process is affected by factors *internal* to the experiment. Endogeneity is highly problematic because accurate estimation of causal relationships requires the cause variable to be at least partially exogenous.

## Threats to Causal Validity

### *Randomized experiments*

The use of randomized experimental research designs ensures that the research subjects in each of the experimental conditions are equal in expectation before the administration of the experimental treatment. However, the use of randomized experimental designs does *not* ensure that the experiment will remain bias-free after randomization. Randomized experiments must be carefully planned and implemented to avoid various biases affecting their results postrandomization. In particular, there are three primary threats (i.e., sources of bias) that must be guarded against for randomized experiments to achieve high levels of causal validity. The first potential threat is *contamination*. Contamination occurs in situations where research subjects assigned to different levels of the experiment (e.g., participants and nonparticipants) come into direct contact or interact in other ways. Contamination occurs when nonparticipants end up receiving the treatment via interactions with participants. For example, if nonparticipants learn ideas/techniques discussed in the experimental treatment, then this knowledge may attenuate the size of the treatment effect because in essence nonparticipants received some of the experiment treatment vicariously. *Cross-overs* are a second potential threat to the causal validity of randomized experimental designs. Cross-overs refer to research subjects assigned to one condition who end up in some other experimental condition. For example, if some nonparticipants end up receiving the treatment because of an error or deliberate actions, then these individuals have "crossed-over." Cross-overs, particularly as their numbers rise, may attenuate the magnitude of the treatment effect and thereby negatively affect the experiment's causal validity. The third potential threat to randomized experimental research designs is *attrition*. Attrition is the loss of research subjects due to factors such as being unable to locate the subjects for follow-up interviews/assessments, subjects declining to participate, death of research subjects, and so forth. Attrition becomes an increasingly potent problem as the length of the tracking period grows. Attrition is problematic in two ways. First, general attrition (i.e., attrition across experimental conditions) undermines external validity, the ability to generalize research findings beyond the sample. Second and more problematic in terms of causal validity is differential attrition (i.e., attrition rates differ markedly between experimental conditions), as differential attrition has the potential to undo the equating of groups accomplished via random assignment.

### *Quasi experiments*

Quasi experiments face a host of issues that threaten the causal validity of findings derived from these designs. The particular threats depend on the specific design features of the quasi experiment. Quasi experiments using one-group designs face the most serious threats to the causal validity of their findings. These threats include *maturation* (i.e., changes due to aging), *regression to the mean* (i.e., the tendency of research subjects who scored unusually high and low scores in initial assessments to regress toward less extreme scores in later assessments), *testing* (i.e., the tendency of research subjects to respond differently in later assessments because they have been sensitized to the behaviors under investigation), and "history" (i.e., external events, besides the intervention, that cause changes in the behaviors under investigation). All of these threats are competing explanations for the results obtained from one-group quasi-experimental research designs. Given the number of threats challenging the causal validity of one-group designs, these designs are the weakest type of experiments.

Two-group quasi-experimental designs generally have fewer and different primary threats to their causal validity in comparison to one-group designs. Briefly, two-group quasi-experimental designs have all of the same threats as randomized experiments and the additional threat of selection bias. As discussed above, the nonrandom assignment of research subjects to experimental conditions leaves open the possibility that research subjects assigned to various levels of the experimental condition differed on observed and unobserved variables before the administration of the experimental treatment. As a result, the variable capturing treatment assignment is potentially endogenous, which is highly problematic because accurate estimation of the treatment effect requires the treatment assignment variable to be exogenous.

## Exogeneity in Nonrandomized Experiments

The use of randomized experiments is not the only means of achieving exogeneity. There are several other research designs/research methods of achieving at least partial exogeneity. These designs/methods are more frequently utilized in fields outside of criminology and are making inroads in criminology. These quasi-experimental designs/methods include natural experiments, regression discontinuity, and instrumental variable estimation. These techniques allow researchers to accurately estimate causal relationships and draw causal inferences in certain situations.

Natural experiments are one means of establishing exogenous variation in a cause variable when researcher-led random assignment is not feasible. A natural experiment is study in which external factors such as natural events, serendipity, or policy changes "assign" research subjects to various experimental conditions of interest. Because the assignment process is external to the research subjects under observation, the assignment process is exogenous, or at least arguably exogenous. This exogenous variation allows researchers to accurately estimate causal relationships.

Natural experiments seem to be increasingly common in the social sciences (see Dunning, 2012). As an example of a natural experiment, Kirk (2009) wished to learn the causal effect of relocating previously incarcerated offenders from their old neighborhoods of residence to new less criminogenic neighborhoods. This is an important theoretical and practical issue because we know that many parolees return to the same criminogenic neighborhoods and social networks that contributed to their involvement in offending in the first place, and therefore we shouldn't be surprised that recidivism is often alarmingly high. While it is not impossible to conduct a randomized experiment on this issue, it would be difficult for a variety of reasons. However, a recent natural event, Hurricane Katrina, forced many parolees who resided in high-crime areas of New Orleans hard hit by the storm to move to other neighborhoods. In essence, Hurricane Katrina exogenously assigned some parolees to new neighborhoods, which made it possible to estimate the causal effect of relocation on recidivism. Kirk found that parolees who moved to a new area were substantially less likely to be reincarcerated within three years of release in comparison to parolees who did not move.

Another research design capable of establishing exogenous variation is the regression discontinuity design (see Murnane & Willett, 2011). The key element of this design is the use of some "forcing variable" that establishes a cut point (or threshold) that assigns research subjects below the cut point to one experimental condition and those above the cut point to another condition. The cut point is used as an exogenous source of variation; research subjects just below and just above the cut point are compared to estimate the causal relationship between the variables of interest. As an example of a regression discontinuity design, Berk and Rauma (1983) assessed the causal effect of providing financial assistance in the form of unemployment insurance to recently released former prison inmates. In order to qualify for the financial assistance former prison inmates had to have made at least $1,500 in the year prior to release; this criterion was used as the forcing variable used to assign former inmates to either the control (no financial assistance) or the treatment (financial assistance) conditions. Berk and Rauma found that financial assistance caused a 13% reduction in recidivism.

Outside of criminology, the most popular means of estimating causal relationships without the use of a random experiment is instrumental variable estimation. The logic of instrumental variable estimation begins by noting that if some part of the variation in some endogenous variable of interest could be established as exogenous, then this part of the variation in the variable of interest could be used to accurately estimate its causal effect on the outcome of interest. An "instrumental variable" can be used to identify exogenous variation. An instrumental variable is one that satisfies two assumptions: (1) it is uncorrelated with the error term of the regression of the outcome variable of interest on the endogenous independent variable of interest; and (2) it is correlated with the endogenous variable of interest. The first assumption means that the instrumental variable can have no effect on the outcome variable except via its indirect effect on the outcome through the endogenous independent variable. This is a strong assumption because the instrumental variable must only be related to the outcome variable through the endogenous independent (causal) variable *and* the instrumental variable cannot be correlated with other factors that affect the outcome variable. If an instrumental variable meeting these criteria can be found, then estimating the causal relationship between the endogenous variable and the outcome of interest is straightforward and can be accomplished using several statistical techniques, of which two-stage least stages is most popular.

Instrumental variable estimation has rarely been applied in criminology. Apel, Bushway, Paternoster, Brame, and Sweeten (2008) is one example of instrumental variable estimation in criminology. Apel and colleagues examine the relationship between hours worked by youth and delinquency. Prior research typically finds that youth who work more hours are more likely to be involved in delinquency; however, number of hours worked is likely to be endogenously related to delinquency, as youth who work more hours are likely to be different from other youth on a host of factors that are also related to delinquency. Apel and colleagues use variation in state child labor laws as an instrumental variable to identify exogenous variation in the number of hours worked. Contrary to prior research, these authors find that the number of hours worked by youth reduces delinquency.

SEE ALSO: Quasi-Experimental Research Methods; Research Methods; Validity and Reliability.

## References


Apel, R., Bushway, S. D., Paternoster, R., Brame, R., & Sweeten, G . (2008). Using state child labor laws to identify the causal effect of youth employment on deviant behavior and academic achievement. *Journal of Quantitative Criminology, 24*, 337–362.

Berk, R. A., & Rauma, D. (1983). Capitalizing on nonrandom assignment to treatments: A regression-discontinuity evaluation of a crime-control program. *Journal of the American Statistical Association, 78*, 21–27.

Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach*. New York: Cambridge University Press.

Farrington, D P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the past two decades? *Journal of Experimental Criminology, 1*, 9–38.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

Kirk, D. S. (2009). A natural experiment on residential change and recidivism: Lessons from Hurricane Katrina. *American Sociological Review, 74*, 484–505.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York: Oxford.

Sampson, R. J., & Laub, J. H. (1993). *Crime in the making: Pathways and turning points through life*. Cambridge, MA: Harvard University Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.


## Further Reading


Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York: Cambridge.