observations can be different, and it is possible to measure *how* different the observations are. Quantitative variables can be either INTERVAL or ratio variables. More and stronger statistical methods are available for the analysis of quantitative variables than for nominal and ordinal variables, REGRESSION analysis being the most important such method.

Examples of quantitative variables are age, income, and number of children in a household. The values of such variables are measured with the help of a unit of measurement. The unit commonly used for age is a year. That way, one person can be 32 years old, and another person can be 25 years old and therefore 7 years younger than the other person. The unit for income is often the dollar, but 1,000 is also used. Number of children is a counting variable, and the common unit is one child.

However, there is nothing about a variable such as age that necessarily makes it a quantitative variable. The complexity of this variable will depend on the way in which it is measured. It could be measured using the values *young, middle aged,* and *old.* Measured this way, age is *not* a quantitative variable; it is an ordinal variable. Also, it could be measured using the values *middle aged* and *not middle aged.* That way it would have to be considered a nominal variable. Thus, in addition to the name of the variable, it is equally necessary to know the way in which the variable is measured.

A less important feature of a quantitative variable is that it can be either DISCRETE or CONTINUOUS. A discrete variable is one for which it is possible to find two values of the variable such that there are no values between the two. Counting variables are discrete; no household can have between two and three children. Similarly, a continuous variable is such that between any two values, it is always possible to find another value. Age, for example, is a continuous variable as we move through time, even though it is often measured as the age at the previous birthday. Sometimes, quantitative variables are called continuous variables.

Quantitative variables are better than nominal or ordinal variables because they can be used for computations such as addition and multiplication. That opens the possibilities for a large variety of statistical computations and analyses. It is possible to find the mean value of a quantitative variable, and it is possible to use the variables for bivariate and multivariate analyses.

—Gudmund R. Iversen

REFERENCES

JMP Introductory Guide. (2000). Cary, NC: SAS Institute.

Moore, D. S., & McCabe, G. P. (1998). *Introduction to the practice of statistics.* New York: W. H. Freeman.

SPSS base 10.0 user's guide. (1999). Chicago: SPSS.

QUARTILE

OBSERVATIONS can be grouped into four equal-sized sets according to their RANK ORDER. Each of the four sets forms a quartile, which is a special case of QUANTILE.

—Tim Futing Liao

See also QUANTILE

QUASI-EXPERIMENT

Quasi-experiments manipulate presumed causes to discover their effects, but the researcher does *not* assign units to conditions randomly. Quasi-experiments are necessary because it is not always possible to randomize. Ethical constraints may preclude withholding effective treatments from needy people based on chance without proper informed consent, those who administer treatment may refuse to honor randomization, or questions about program effects may arise after a TREATMENT was already implemented so that randomization is impossible. So, quasi-experiments use a combination of design features, practical logic, and statistical analysis to show that the treatment may be a plausible cause of the effect. The resulting causal inferences are often more ambiguous than is the case with randomized experiments. Nonrandomized experiment is synonymous with quasi-experiment, and observational study and nonexperimental design often include quasi-experiments as a subset.

KINDS OF QUASI-EXPERIMENTAL DESIGNS

Quasi-experimental designs include, but are not limited to, (a) nonequivalent control group designs, in which the outcomes of those exposed to two or more conditions are studied but the experimenter does not control assignment to conditions; (b) INTERRUPTED TIME-SERIES DESIGNS, in which many consecutive observations over time (prototypically 100) are

available on an outcome, and treatment is introduced in the midst of those observations to demonstrate its impact on the outcome through a discontinuity in the time series after treatment; (c) regression discontinuity designs, in which the experimenter uses a cutoff score on a measured variable to determine eligibility for treatment, and an effect is observed if the regression line of the assignment variable on outcome for the treatment group is discontinuous from that of the comparison group; and (d) single-case designs, in which one participant is observed repeatedly over time (usually on fewer occasions than in the time series) while the scheduling and dose of treatment are manipulated to demonstrate that treatment controls outcome.

In such designs, treatment is manipulated, and outcome is then observed. Two other classes of designs are sometimes included as quasi-experiments, even though the cause is not manipulated. In (e) case control designs, a group with an outcome of interest is compared to a group without that outcome to see if they differ retrospectively in exposure to possible causes; and in (f) correlational designs, observations on possible treatments and outcomes are observed simultaneously, often with a survey, to see if they are related. Because these designs do not ensure that cause precedes effect, as it must logically do, they usually yield more equivocal causal inferences.

HISTORICAL DEVELOPMENT

Most experiments conducted prior to the 1920s were quasi-experiments. For example, Lind (1753) described a quasi-experimental comparison of six medical treatments for scurvy; around 1850, epidemiologists used case control methods to identify contaminated water supplies as the cause of cholera in London; and in 1898, Triplett used a nonequivalent control group design to show that the presence of an audience and competitors improved the performance of bicyclists.

In 1963, Campbell and Stanley coined the term *quasi-experiment* to describe this class of designs. Campbell and his colleagues (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002) extended the theory and practice of these designs in three ways. First, they described a larger number of these designs. For example, some quasi-experimental designs are inherently longitudinal (e.g., time series, single-case designs), observing participants over time; but other designs can be made longitudinal by adding more observations before or after treatment. Similarly, more than one treatment or control group can be used, and the designs can be combined, as when adding a nonequivalent control group to a time series.

Second, Campbell developed a logic to evaluate the quality of causal inferences resulting from quasi-experimental designs—a VALIDITY typology elaborated in Cook and Campbell (1979) and Shadish et al. (2002). The typology includes four validity types and threats to validity for each type. Threats are common reasons why researchers may be wrong about the causal inferences they draw. Statistical conclusion validity concerns inferences about how much presumed cause and effect covary; an example of a threat is low STATISTICAL POWER. INTERNAL VALIDITY concerns inferences that observed covariation is due to the treatment causing the outcome; a sample threat is history (extraneous events that could also cause the effect). CONSTRUCT VALIDITY concerns inferences about higher-order constructs that research operations represent; a sample threat is EXPERIMENTER EXPECTANCY EFFECTS, whereby participants react to what they believe the experimenter wants to observe rather than to the intended treatment. EXTERNAL VALIDITY concerns inferences about whether the cause-effect relationship holds over variation in people, settings, treatment variables, and measurement variables; threats include interactions of the treatment with other features of the design that produce unique effects that would not be observed otherwise.

Third, Campbell addressed threats to validity using design features—things a researcher can manipulate to prevent a threat from occurring or to diagnose its presence and impact on study results (Table 1). For example, suppose maturation (normal development over time) is an anticipated threat to validity because it could cause a pretest-posttest change like that attributed to treatment. The inclusion of several consecutive pretests before treatment can indicate whether the rate of maturation before treatment is similar to the rate of change during and after treatment. If so, maturation is a threat. All quasi-experiments are combinations of these design features, chosen to diagnose or rule out threats to validity in a particular context. Campbell was skeptical about adjusting threats statistically after they have already occurred because statistical adjustments require making assumptions that are usually dubious or impossible to test.

Other scholars during this time were also interested in causal inferences in quasi-experiments, particularly

Table 1 Design Elements Used in Constructing Quasi-Experiments

Assignment (Control of assignment strategies to increase group comparability)

- *Cutoff-Based Assignment.* Controlled assignment to conditions based on one or more fully measured covariates. This yields an unbiased effect estimate.
- *Other Nonrandom Assignment.* Various forms of "haphazard" assignment that sometimes approximate randomization (e.g., alternating assignment in a two-condition quasi-experiment whereby every other unit is assigned to one condition, etc.).
- *Matching and Stratifying.* Efforts to create groups equivalent on observed covariates in ways that are stable, do not lead to regression artifacts, and are correlated with the outcome.

Measurement (Use of measures to learn whether threats to causal inference actually operate)

Posttest Observations

- *Nonequivalent Dependent Variables.* Measures that are not sensitive to the causal forces of the treatment, but *are* sensitive to all or most of the confounding causal forces that might lead to false conclusions about treatment effects (if such measures show no effect, but the outcome measures do show an effect, the causal inference is bolstered because it is less likely to be due to the confounds).
- *Multiple Substantive Posttests.* Used to assess whether the treatment affects a complex pattern of theoretically predicted outcomes.

Pretest Observations

- *Single Pretest.* A pretreatment measure on the outcome variable, useful to help diagnose selection bias.
- *Retrospective Pretest.* Reconstructed pretests when actual pretests are not feasible—by itself, a very weak design feature, but sometimes better than nothing.
- *Proxy Pretest.* When a true pretest is not feasible, a pretest on a variable correlated with the outcome—-often weak by itself.
- *Multiple Pretest Time Points on the Outcome.* Helps reveal pretreatment trends or regression artifacts that might complicate causal inference.
- *Pretests on Independent Samples.* When a pretest is not feasible on the treated sample, one is obtained from a randomly equivalent sample.
- *Complex Predictions Such as Predicted Interaction.* Successfully predicted interactions lend support to causal inference because alternative explanations become less plausible.
- *Measurement of Threats to Internal Validity.* Help diagnose the presence of specific threats to the inference that A caused B, such as whether units actively sought out additional treatments outside the experiment.

Comparison Groups [Selecting comparisons that are "less nonequivalent" or that bracket the treatment group at the pretest(s)]

- *Single Nonequivalent Groups.* Compared to studies without control groups, using a nonequivalent control group helps identify many plausible threats to validity.
- *Multiple Nonequivalent Groups.* Serve several functions. For instance, groups are selected that are as similar as possible to the treated group, but at least one outperforms it initially and at least one underperforms it, thus bracketing the treated group.
- *Cohorts.* Comparison groups chosen from the same institution in a different cycle (e.g., sibling controls in families or last year's students in schools).
- *Internal (vs. External) Controls.* Plausibly chosen from within the same population (e.g., within the same school rather than from a different school).

Treatment (Manipulations of the treatment to demonstrate that treatment variability affects outcome variability)

- *Removed Treatments.* Shows that an effect diminishes if treatment is removed.
- *Repeated Treatments.* Reintroduces treatments after they have been removed from some group—common in laboratory sciences or where treatments have short-term effects.
- *Switching Replications.* Reverses treatment and control group roles so that one group is the control while the other receives treatment, but the controls receive treatment later while the original treatment group receives no further treatment or has treatment removed.
- *Reversed Treatments.* Provides a conceptually similar treatment that reverses an effect (e.g., reducing access to a computer for some students but increasing access for others).
- *Dosage Variation.* Demonstrates that outcome responds systematically to different levels of treatment.

William G. Cochran in statistics, James J. Heckman in economics, and Sir Austin Bradford Hill in epidemiology. However, Campbell's work was unique for its extensive emphasis on design rather than statistical analysis, its theory of how to evaluate causal inferences, and its sustained development of quasi-experimental theory and method over four decades.

EXAMPLES

In 1966, the Canadian province of Ontario initiated a formal program for the screening and treatment of infants born with phenylketonuria (PKU) to prevent retardation. After the start of the program, 44 infants born with PKU experienced no retardation, and three did. Of these, two were missed by the screening program (Webb et al., 1973). Statistics from prior years suggested a much higher rate of retardation attributable to PKU. Although this study lacked a control group, the authors concluded that the program successfully treated PKU infants. Based on such results, this program was widely adopted in Canada and the United States. This study was a pretest-posttest quasi-experiment with no control group.

In July 1982, Arizona implemented legislation mandating severe penalties for driving while intoxicated; a comparison of monthly results from January 1976 to July 1982 (the control condition) with monthly totals between July 1982 and May 1984 (the treatment condition) found a decrease in traffic fatalities after the new law was implemented. A similar finding occurred in monthly data trends in San Diego after January 1982, when that city implemented a California law that also penalized intoxicated drivers. In a control time series in El Paso, Texas, a city that had no relevant change in its driving laws during this period, monthly fatality trends showed no comparable changes during the months of either January or July 1982. The changes in trends over time in both San Diego and Arizona, compared to the absence of similar trends in El Paso, suggest that the new laws reduced fatalities (West, Hepworth, McCall, & Reich, 1989). This study was an interrupted time-series quasi-experiment.

STATISTICS AND QUASI-EXPERIMENTAL DESIGN

Statisticians such as Paul Holland, Paul Rosenbaum, and Donald Rubin emphasize the need to measure what would have happened to treatment participants without treatment (the counterfactual) and focus on statistics that can improve estimates of the counterfactual without randomization. A central method uses PROPENSITY SCORES, a predicted probability of group membership obtained from LOGISTIC REGRESSION of actual group membership on predictors of outcome or of how participants got into treatment. MATCHING, stratifying, or covarying on the propensity score can balance nonequivalent groups on those predictors, but those methods cannot balance groups for unobserved variables, so hidden bias may remain. Hence, these statisticians have developed sensitivity analyses to measure how much hidden bias would be necessary to change an effect in important ways.

Economists such as James Heckman and his colleagues have pursued another development, SELECTION BIAS modeling, which aims to remove hidden bias from quasi-experiments by modeling the selection process. Unfortunately, these models have not been very successful in matching results from randomized experiments. Most recently, economists have improved results by combining selection bias models with propensity scores. This topic continues to develop.

A third development is the use of STRUCTURAL EQUATION MODELING (SEM) to study causal relationships in quasi-experiments; this effort has also been only partly successful (Bollen, 1989). The capacity of SEM to model latent variables can sometimes reduce bias caused by unreliability of measurement, but its capacity to generate unbiased effect estimates is hamstrung by the same lack of knowledge of selection that thwarts selection bias models. A related but newer literature on causality is promising, using directed graphs to help understand the issues (Pearl, 2000).

CONCLUSION

Quasi-experiments are rarely able to provide the confidence about causal inference that randomized experiments can provide, and overreliance on quasi-experiments in some areas is a serious problem. Nevertheless, three important factors have created conditions to improve quasi-experiments. First, extensive practical experience with quasi-experimental designs has provided a database from which we can conduct empirical studies of the theory (Shadish, 2000). Second, after decades of focus on randomized designs, statisticians and economists have turned their attention to improving quasi-experimental designs. Third,

the computer revolution provided both theorists and practitioners with increased capacity to invent and use more sophisticated and computationally intense methods for improving quasi-experiments.

—William R. Shadish and M. H. Clark

REFERENCES

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand-McNally.

Cochran, W. G. (1965). The planning of observational studies in human populations. *Journal of the Royal Statistical Society, Series A, 128,* 134–155.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand-McNally.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47,* 153–161.

Hill, A. B. (1953). Observation and experiment. *New England Journal of Medicine, 248,* 995–1001.

Lind, J. (1753). *A treatise of the scurvy: Of three parts containing an inquiry into the nature, causes and cure of that disease.* Edinburgh, UK: Sands, Murray and Cochran.

Pearl, J. (2000). *Causality, models, reasoning and inference.* New York: Cambridge University Press.

Rosenbaum, P. R. (1995). *Observational studies.* New York: Springer-Verlag.

Schlesselman, J. J. (1982). *Case-control studies: Design, conduct, analysis.* New York: Oxford University Press.

Shadish, W. R. (2000). The empirical program of quasi-experimentation. In L. Bickman (Ed.), *Validity and social experimentation: Donald Campbell's legacy* (pp. 13–35). Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology, 9,* 507–533.

Webb, J. F., Khazen, R. S., Hanley, W. B., Partington, M. S., Percy, W. J. L., & Rathborn, J. C. (1973). PKU screening—is it worth it? *Canadian Medical Association Journal, 108,* 328–329.

West, S. G., Hepworth, J. T., McCall, M. A., & Reich, J. W. (1989). An evaluation of Arizona's July 1982 drunk driving law: Effects on the City of Phoenix. *Journal of Applied Social Psychology, 19,* 1212–1237.

Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology, 25,* 659–707.

QUESTIONNAIRE

All scientists use "tools" to measure the phenomena of interest to their disciplines, and social scientists are no exception. Thus, just as a telescope serves astronomers and an electron microscope serves microbiologists, the questionnaire serves many social scientists as their primary measurement tool. The challenge to these researchers is to calibrate a questionnaire that will gather data as accurately and efficiently as possible.

From a "total error" perspective, a poorly constructed questionnaire can contribute BIAS and VARIANCE to the data that are gathered (cf. Groves, 1989). In particular, poorly worded, poorly ordered, and/or poorly formatted questions can lead to significant measurement error in the form of both bias and variance, and/or can lead to significant item NONRESPONSE error in the form of bias.

One of the defining characteristics of a questionnaire is whether it is for a SELF-ADMINISTERED survey, as in a mail survey or an INTERNET SURVEY, or whether it is to be *interviewer-administered,* as in a telephone survey or in-person survey. If the questionnaire is self-administered, then the "end user" is the respondent. In these cases, all instructions required to complete the questionnaire accurately must be stated clearly and explicitly within the questionnaire itself. Or, in the case of computerized self-administered questionnaires, instructions can be presented via audio recordings. There are myriad other design features that are affected by whether the questionnaire is self-administered or interviewer-administered; for detailed information, see Dillman (2000).

In creating a particular questionnaire *item,* there are four structural factors to consider: (a) the *question stem* wording; (b) whether the *response option(s)* will be open-end or closed-end; and, if closed-end, whether they are (c) forced-choice and/or (d) balanced. In wording the question stem (i.e., the interrogative being asked), use as few and as simple words possible to convey the meaning of the construct being measured. In operationalizing the construct through this wording, the researcher also must keep in mind the educational level of the respondents. Whenever a questionnaire is self-administered, respondents' literacy must be considered.