22

SINGLE-CASE RESEARCH DESIGNS

MATTHEW K. NOCK

BETHANY D. MICHEL

VALERIE I. PHOTOS

Single-case research designs are a diverse and powerful set of procedures useful for demonstrating causal relations among clinical phenomena. Although such designs are flexible, efficient, and have been used to make key advances since the earliest days of psychological science, they are currently extremely underused by psychological scientists and clinicians. We review the historical and theoretical foundations of single-case research designs, including case studies, quasi-experimental designs, and experimental designs used in single-case or small-group studies. We also summarize the methodological requirements, primary design options, and challenges and limitations of each approach. Recent examples of each design are provided to illustrate the usefulness of these valuable—albeit somewhat forsaken—research methods for scientific and clinical advancement.

HISTORICAL AND THEORETICAL FOUNDATIONS

The goal of science is to establish and advance knowledge. Scientists use careful and systematic procedures for observing relations among phenomena (that is, variables), and in doing so attempt to rule out alternative explanations for the observed relations. There are many different research arrangements and procedures that can be used to draw valid inferences about the relations between variables of interest—as evidenced by the thickness of this volume. This chapter reviews *single-case research designs,* which refers to those in which the phenomena of interest are studied using a single subject or a small group of research subjects (also referred to as *N*-of-1 designs or single-subject research). Single-case research designs have a long and distinguished history in psychological science, which demonstrates the enduring value of such approaches.

Psychological science grew out of earlier work in philosophy and physiology, and from its first days as a new science, psychology relied on in-depth empirical studies that used single subjects or small groups of subjects. For instance, many of the founders of psychological science, including Wundt, Ebbinghaus, and Pavlov, conducted very careful experiments of single subjects or small groups of subjects to make classic discoveries in the study of perception, learning and memory, and the laws of conditioning, respectively (Schultz & Schultz, 1992). As an interesting example, Ebbinghaus conducted ground-breaking work on learning and memory over the course of five years using only

one research subject—himself (Ebbinghaus, 1885/1913).

The in-depth study of the single case was also the primary method of investigation in the earliest days of modern clinical psychology. For instance, the most influential figures in psychoanalytic psychology used naturalistic case studies to convey their ideas about the etiologies, assessment, and treatment of psychopathology (Freud & Breuer, 1895). Although working from a very different theoretical and methodological orientation, the founders of behavioral psychology also used single-case research designs to inform research and practice in psychology and psychopathology (Skinner, 1938; Watson, 1925).

If single-case research designs have led to such broad and significant advances since the earliest days of psychological science, why has their use declined so markedly over the past several decades? While there are many possible explanations, the most compelling may be the development of increasingly sophisticated methods for collecting and analyzing large amounts of data. Indeed, with the development of statistical techniques appropriate for use with large samples (e.g., *t* test, analysis of variance, correlation), as well as the creation of computer programs that facilitate the evaluation of these data, researchers began to favor large sample designs that focused on statistical evaluation over single-case designs that relied primarily on experimentation and visual inspection of data. This focus continues to the present day, and research studies that examine aggregate data from large groups of subjects continue to compose the vast majority of studies reported in the mainstream, high-impact clinical psychology journals such as the *Journal of Consulting and Clinical Psychology* and the *Journal of Abnormal Psychology,* while single-case research designs represent the majority of reports published in specialty journals such as the *Journal of Applied Behavior Analysis* and the *Journal of the Experimental Analysis of Behavior.*

Although the advances that have resulted from the proliferation of large sample studies is undeniable, single-case research designs continue to be underutilized by researchers and clinicians alike. We believe efforts to reincorporate such designs into the clinical research armamentarium will facilitate significant advances in psychological science.

In this chapter, we provide a brief review of the three types of single-case research designs most commonly used by clinical researchers: the case study, quasi-experimental designs, and experimental designs. It is probably more appropriate to think of these three design options not as distinct types but more accurately as points on a continuum. These designs all share common features, such as the examination of a single or small group of subjects; however, they are characterized by increasing levels of scientific rigor, and, as an effect, by the strength of the inferences that can be made with their use. For each design, we describe the basic characteristics, primary research design options, recent advances, and challenges and limitations. Our depth of coverage of this material is limited by the length of this chapter; the reader interested in more extensive coverage of each should consult several excellent sources on these topics (Barlow, Hayes, & Nelson, 1984; Barlow & Hersen, 1984; Franklin, Allison, & Gorman, 1997; Kazdin, 1982, 2003).

Case Studies

Case studies refer to a broad array of approaches most often used for observing an individual and reporting on their interactions with variables of interest (e.g., life events, psychological intervention, and so on). Although broad, case studies generally are similar in their inclusion of several primary characteristics (Kazdin, 2003). First, case studies most often involve the study of an individual. This may be an individual person, family, group, or classroom; however, the focus is the study of a single unit. Second, data collected in case studies are detailed, qualitative, and anecdotal, but are rarely systematically or quantitatively measured. Third, there is a strong focus on the unique aspects of the case, allowing the researcher to note complexities arising from the distinctive history and influences specific to that individual. Finally, data typically are collected retrospectively and no experimental controls are applied.

Although observational in nature, case studies offer several significant advantages over more commonly used group-based studies. For instance, because case studies do not require control conditions or comparison groups, they can be easily incorporated into routine clinical work without disrupting the natural pace of treatment. In this way, case studies offer clinicians both a useful research tool as well as a source of data directly relevant to clinical practice. Case studies also are valuable in the study of rare phenomena. This is

well illustrated in Oliver Sacks' study of Dr. P, who famously misperceived his wife's head as his own hat (Sacks, 1990). In cases where impairments in functioning are severe and important but occur so infrequently as to preclude subject recruitment for large sample studies, case studies can provide critical information about the phenomena of interest. Similarly, the absence of methodological restrictions or structure imposed by measurement techniques provides an opportunity to observe the behavior of interest as it naturally occurs, which may be useful for generating hypotheses that can subsequently be tested using empirical methods.

Case studies are also useful for the development and evaluation of novel assessment and treatment techniques. Clinicians and researchers share an interest in developing and disseminating effective assessment and treatment methods. Doing so requires considerable time and resources, including multiple revisions to existing methods and relatively long delays in obtaining funding and publishing the findings of clinical trials. Case studies provide an incredibly efficient and flexible tool in the assessment and treatment development and evaluation process. As an example, case studies can be effectively used to modify multiple iterations of a new treatment and to collect initial feasibility or efficacy data for such a treatment—which typically is a necessary step in obtaining funding for a larger treatment study (see Moras, Telfer, & Barlow, 1993).

Although case studies can provide valuable information, as commonly used they lack the methodological requirements to draw valid inferences about the relations among variables (Kazdin, 1981). Indeed, case studies typically do not include design features such as objective assessment, systematic data collection or analysis, specified manipulation of the independent variable, or replication of treatment effects. Therefore, one is limited in the validity of the inferences that can be drawn from such evaluations. Fortunately, such design features can be easily incorporated into single-case research designs in order to strengthen the conclusions drawn from such cases.

Single-Case Quasi-Experimental Designs

Scientists learn about the world in general, and about abnormal and clinical psychology in the present case, by carefully observing and measuring variables and the relations among them (that is, *observational research*) and in some cases by systematically manipulating certain variables and measuring the effect on other variables (that is, *experimental research*). The purpose of the research enterprise is to be able to arrive at truthful conclusions regarding the relations among variables—the more careful the measurement, control, and manipulation over the variables of interest, the more valid the conclusions that can be drawn. All research exists somewhere along a continuum regarding the amount of precision and control involved.

The scientists' ultimate goal is to rule out threats to the validity of their conclusions about the relations between variables by systematically controlling different, and ideally all, facets of their investigation. With increasing levels of control over the observation and manipulation of variables come increasing degrees of confidence in the conclusions that can be drawn. That is, if you control all of the factors influencing your variables, you can make valid conclusions about why they changed. In case studies, there is very little control on assessment and manipulation of variables. On the other end of the continuum are true experimental studies. Such studies are characterized by systematic measurement and manipulation of individual variables in order to demonstrate that with all other variables held constant, controlled variation in the independent variable is associated with variation in the dependent variable—strengthening the argument for a causal relation between the two.

In instances in which the level of control and manipulation required in a true experimental design is not possible, researchers may use a *quasi-experimental design,* in which some elements of an experimental design are used in order to increase the validity of the inferences drawn from such an investigation. In such cases, certain design elements can be added that allow the experimenter to arrive at stronger conclusions than would be possible using only a case study design.

Design Essentials

Repeated Assessment

Perhaps the most essential methodological requirement for single-case experimental designs is repeated assessment of the dependent variable

across all phases of the study. The need for repeated assessment stems from the fact that rather than evaluating overall between-group differences on the dependent variable (as in most between-group designs), one evaluates patterns of within-subject changes as the individual moves temporally through different experimental conditions.

Measurement of the dependent variable(s) should use reliable and valid measures, and should occur *early, frequently,* and *consistently.* Assessing early is important in order to establish a baseline, or preintervention level of responding, which is characteristic of many single-case quasi-experimental designs and all single-case experimental designs. Regardless of whether baseline data collection is possible, assessment should commence as early in the study as possible. Assessing frequently is important in order to capture change as it is occurring during the intervention. The actual timing and frequency of assessments will vary depending on the phenomenon under evaluation, the assessment method used, and the hypotheses about how quickly change is likely to occur. For instance, assessment may include constant measurement over a single session, behavioral sampling over discrete time intervals, once-per-day measurement, or any of a range of other assessment options (see Nock & Kurtz, 2005). Finally, using consistent assessment methods across assessment occasions is necessary in order to rule out the possibility that behavioral changes are due to changes in assessment procedures.

In making decisions about assessment methods, the experimenter must balance using reliable, valid, and comprehensive methods with what is feasible in a given research or clinical context. Devising appropriate assessment methods may require flexibility and creativity, and in clinical settings the experimenter should include information and data that is already being collected whenever possible (e.g., diary cards completed during the week, reports from teachers or parents, and so on). Indeed, adding unnecessary measures may increase subject burden and decrease the likelihood of adherence to the experimental protocol.

Specified Intervention or Manipulation

Another very important methodological requirement for single-case experimental designs is the use of a specified intervention or experimental manipulation. The intervention used serves as the independent variable, and in order to make clear and interpretable statements about the relation between the intervention and the behavioral outcome, one must know what the intervention actually was. Indeed, careful control over and record of the intervention is necessary in order to demonstrate construct validity and to replicate the observed relations. It is important to note that the intervention need not be behavioral in orientation (that is, it can be composed of cognitive, psychodynamic, or other methods). The only requirement is that it be well specified. The importance of carefully operationalizing the independent variable and replicating one's findings has been covered in detail in other sources, and their importance should be clear to the reader.

Example of a Single-Case Quasi-Experimental Design

As an example, one recent study used a single-case quasi-experimental design to evaluate the effectiveness of a behavioral treatment for panic disorder in a 10-year-old boy (Nock, Goldman, Wang, & Albano, 2004). In this study, the boy, Michael, presented to an outpatient psychiatric clinic with frequent and repeated panic attacks, which were unsuccessfully treated in a previous play therapy. Michael was treated using a modified version of Panic Control Treatment, a manualized treatment developed for and evaluated for adults (Barlow & Craske, 2000). Thus, treatment was very well specified in advance and was carefully administered over the course of treatment. In addition to using semi-structured clinical interviews at pre- and posttreatment to evaluate symptoms of panic disorder, Michael and his mother both completed and turned in daily logs of the number of panic attacks experienced as well as his overall level of global anxiety each day. Michael experienced between one and five panic attacks per week for the first 10 weeks of treatment, which decreased to zero by the 11th week and remained at this level for the rest of the treatment period (except for one panic attack during the 22nd week of treatment) and when assessed at six-month follow-up (Nock et al., 2004).

The use of repeated assessment and a well-specified intervention in this study provided useful information about the intervention used as well as the timing and process of change over

the course of treatment. However, given that there was no control (that is, no intervention) condition, it is possible that other factors could have caused the observed behavior change. The ability to make definitive causal conclusions about the relations between variables is possible only with the use of a true experimental design.

Single-Case Experimental Designs

Experimental research as currently practiced in psychological science most often involves randomly assigning a relatively large number of subjects to different conditions (that is, levels of the independent variable) and observing the effects on the dependent variable of interest. However, random assignment and large numbers of subjects are not necessary to demonstrate causal relations. All that is needed is controlled variation of the independent variable and measurement of the effects of this variation. Single-case experimental designs use all of the previously mentioned design features and (very important) incorporate controlled variation techniques, allowing the experimenter to examine causal relations between intervention and outcome.

Design Essentials

Multiple Phases and Conditions

Inferences about causal effects in single-case experimental designs are drawn by analyzing differences between experimental conditions and control conditions, just as in most between-group designs. The primary difference is that in between-group studies individuals are assigned to *either* an experimental *or* control condition; however, in single-case experimental designs each individual participates in all conditions. The number and ordering of conditions varies depending on the type of design employed (more on this later). The vital element is the inclusion of multiple conditions (that is, >1) for each subject.

Single-case experimental designs typically begin with a baseline condition in order to evaluate performance or functioning in the absence of the intervention. The baseline period, and each subsequent phase, should continue until stability in the dependent variable is observed. The presence of stable responding increases the experimenter's ability to make valid inferences about the effects of the manipulation. That is, stable performance will provide data about both the observed and expected (that is, predicted) behavior, and as in all investigations in which inference is the goal, we are here interested in contrasting observed and expected effects. For instance, as shown in Figure 22.1, data collected during the baseline period (A_1) in which there is no intervention demonstrates a certain level of performance. If the intervention is applied and it is completely ineffective, performance will continue at this level (dotted line in B_1).

The effectiveness of interventions is tested by evaluating the presence and degree of change in the dependent variable(s) that occurs with the introduction of clinician-determined changes in conditions or phases. If all extraneous (that is, nonexperimental) variables are constant and the only variable that changes is the intervention, then one can conclude that any concurrent change in the dependent variable was caused by this change. For instance, in Figure 22.1, when the intervention is introduced (B_1), the dependent variable is observed at a rate that differs markedly from expected, suggesting that the intervention is responsible for this change (more on data analysis later). Of course, it is rarely possible to completely control all extraneous variables, so it is still plausible that some variable other than the intervention (e.g., change in relationships, Yankees winning World Series, or such) caused the observed change. Such threats can be ruled out with subsequent replications of the experimental effects.

Replication of Experimental Effects

Replication is an integral part of scientific advancement and is necessary for garnering support for the operation of causal relations (Kazdin & Nock, 2003). Within the context of single-case experimental designs, replication of experimental effects can and should occur both within and between studies. Each of the single-case experimental design options described following contains methods for such replication—which is one of the great strengths of such approaches.

Primary Research Designs

Researchers and clinicians can use the basic principles and methodological elements described to create a wide variety of single-case

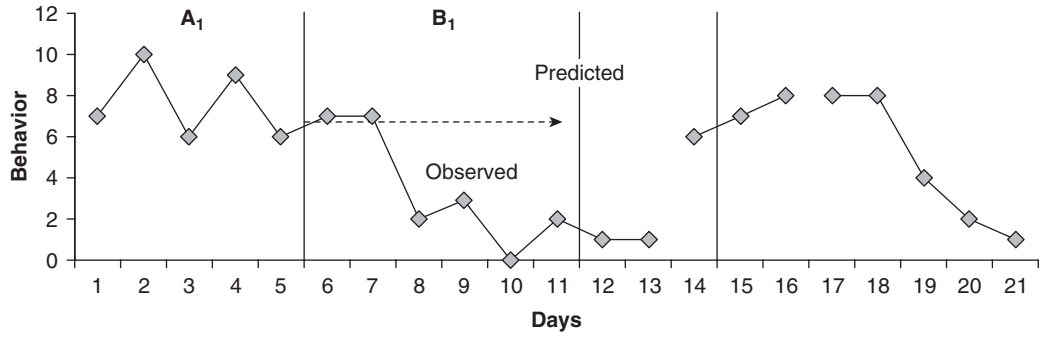

Figure 22.1 Hypothetical Data for Frequency of Behavior Across a Reversal Design

experimental designs. The number of different designs, and variations on these designs, are too numerous to review here. We review three of the most commonly employed single-case experimental designs next. These designs all use repeated assessment, specified interventions, and multiple phases, but differ in the number and ordering of phases as well as in the manner in which they demonstrate causal relations.

Reversal Designs

Reversal designs, also known as *ABAB designs,* are those in which the subject's behavior is first assessed during a baseline phase (A_1), then during an experimental phase in which the intervention/manipulation is administered (B_1), followed by a return to the baseline phase (A_2) during which the behavior is expected to reverse to its baseline level, and so on (see Figure 22.1). Assessment during the baseline phase serves as a control condition, which is necessary to establish the level of the behavior in the absence of the intervention and to demonstrate that any observed behavior change is not merely due to factors such as therapeutic contact or the delivery of assessment measures. The demonstration that a stable pattern of behavior changes when *and only when* the experimental phase begins suggests that the intervention is responsible for this behavior change. Of course, as mentioned in the discussion of the quasi-experimental design, it is possible that other factors are responsible for this change, such as history, maturation, or statistical regression. These threats to the internal validity of the study are ruled out by the reversal or return to baseline condition. If the subject's behavior returns to or approaches the baseline level when the intervention is withdrawn, this provides further evidence for a causal relation between the intervention and the behavior change. It is common after an ABA series to reapply the intervention (that is, ABAB) in order to replicate the treatment effect, as well as for obvious clinical and ethical reasons.

Considerations. In cases in which the specified intervention was not effective (that is, did not lead to behavior change) or was not as effective as expected, the researcher may wish to test the effect of adding an additional treatment component or a modified version of the intervention. This is a common variation of the reversal design. In such cases, additional experimental phases are denoted by consecutive letters (e.g., ABAC design) or combinations of letters in cases in which different treatment components are combined (e.g., A, B, A, B, C). Of course, when such designs are used one must be cognizant of potential treatment sequencing effects, in which the effects are due to a particular sequence of treatment administration, rather than to the most recently administered intervention. The use of additional subjects and the alteration of the sequence of phases can be used to evaluate the influence of sequencing effects.

The demonstration that the presence versus absence of an intervention can be used to repeatedly change the occurrence of a given behavior provides a convincing argument for causality. However, even if the intervention caused a behavior change, there are instances in which it is difficult to cause the behavior to reverse to baseline levels. For instance, if treatment effects

were caused or maintained by a process that is not easily removed, such as the development of problem-solving skills, it will be difficult to demonstrate a reversal. There are also instances in which it is undesirable to cause the behavior to return to baseline levels. For instance, if an intervention reduced self-injurious or aggressive behavior toward others, it may not be clinically or ethically advantageous to bring back such behaviors in order to strengthen support for causal relations. In such instances, other designs may be more appropriate to evaluate causal relations between intervention and behavior change.

Multiple-Baseline Designs

Multiple-baseline designs are those in which the experimental condition is introduced in temporal sequence to different behaviors, settings, or subjects. In such designs, there is typically only one baseline (that is, no reversal) and one intervention condition. The power of such designs comes from demonstrating that change occurs when, and only when, the intervention is directed at the behavior, setting, or subject in question. As with reversal designs, multiple-baseline designs begin with a baseline phase that continues until behavioral stability is demonstrated, at which time the intervention condition begins. As mentioned previously, a change in behavior that occurs only when the intervention is introduced suggests that the intervention caused the change; however, the influence of other factors (e.g., history, maturation, and so on) must be ruled out in order to increase the validity of this claim. Rather than using a return to baseline, multiple baseline designs replicate the intervention—behavior change relation in temporal sequence across different behaviors, settings, or individuals. The temporal sequencing element is vital in order to rule out the likelihood that extraneous factors could account for the observed behavior change.

As an example, one recent study used a multiple-baseline design to evaluate the effectiveness of a behavioral intervention for food refusal in a 4-year-old boy (Nock, 2002). Since he was 7 months old, the boy, Antonio, refused to consume virtually all solid foods or liquids, swallowing only water, oatmeal, soft baby food, and protein drinks, a behavior that could lead to serious health problems (Nock, 2003). In addition to administering a semi-structured interview and a behavior avoidance test at pre- and post-treatment, assessment included a daily log, maintained by his parents and converted in number of servings per day by the clinician, of all of the foods and liquids Antonio consumed throughout the course of the intervention. The intervention involved using modeling and contingency management techniques to gradually expose Antonio to eating an increasing variety of foods and liquids (see Figure 22.2). After an initial two-week baseline period, the intervention first focused on increasing Antonio's consumption of liquids (Figure 22.2A). Once his intake of liquids was increased and stable, the intervention then was applied to the consumption of soft foods (Figure 22.2B), followed by hard foods (Figure 22.2C), and finally chewy foods (Figure 22.2D). The intervention occurred in weekly clinic-based sessions, where Antonio's father learned to administer the intervention, which he did each day at home over the course of the study. As shown in Figure 22.2, visual inspection of the data demonstrates that the consumption of each group of food or liquid increased when *and only when* the intervention was applied to that group. This pattern of results strongly suggests that the intervention caused the behavior change rather than alternative factors such as history, maturation, relationship with the therapist, and the like.

Considerations. Among the many strengths of multiple-baseline designs is that they are very flexible and can demonstrate causality by replicating treatment effects across different behaviors (as in Antonio's case), different settings (e.g., school, home, and work), or different individuals (e.g., different patients, classmates, or group members). Moreover, multiple-baseline designs do not require the withdrawal of effective interventions, thus avoiding some of the clinical and ethical concerns often raised against the use of reversal designs. One potential limitation of multiple-baseline designs is that a convincing demonstration of causality requires that change does not occur across behaviors, settings, or individuals until the intervention is applied; therefore, interventions that lead to changes that generalize across all areas will limit the strength of the inferences that can be drawn. Such a state of affairs is of course very desirable from a clinical perspective, but it leads to some limitations from a research perspective.

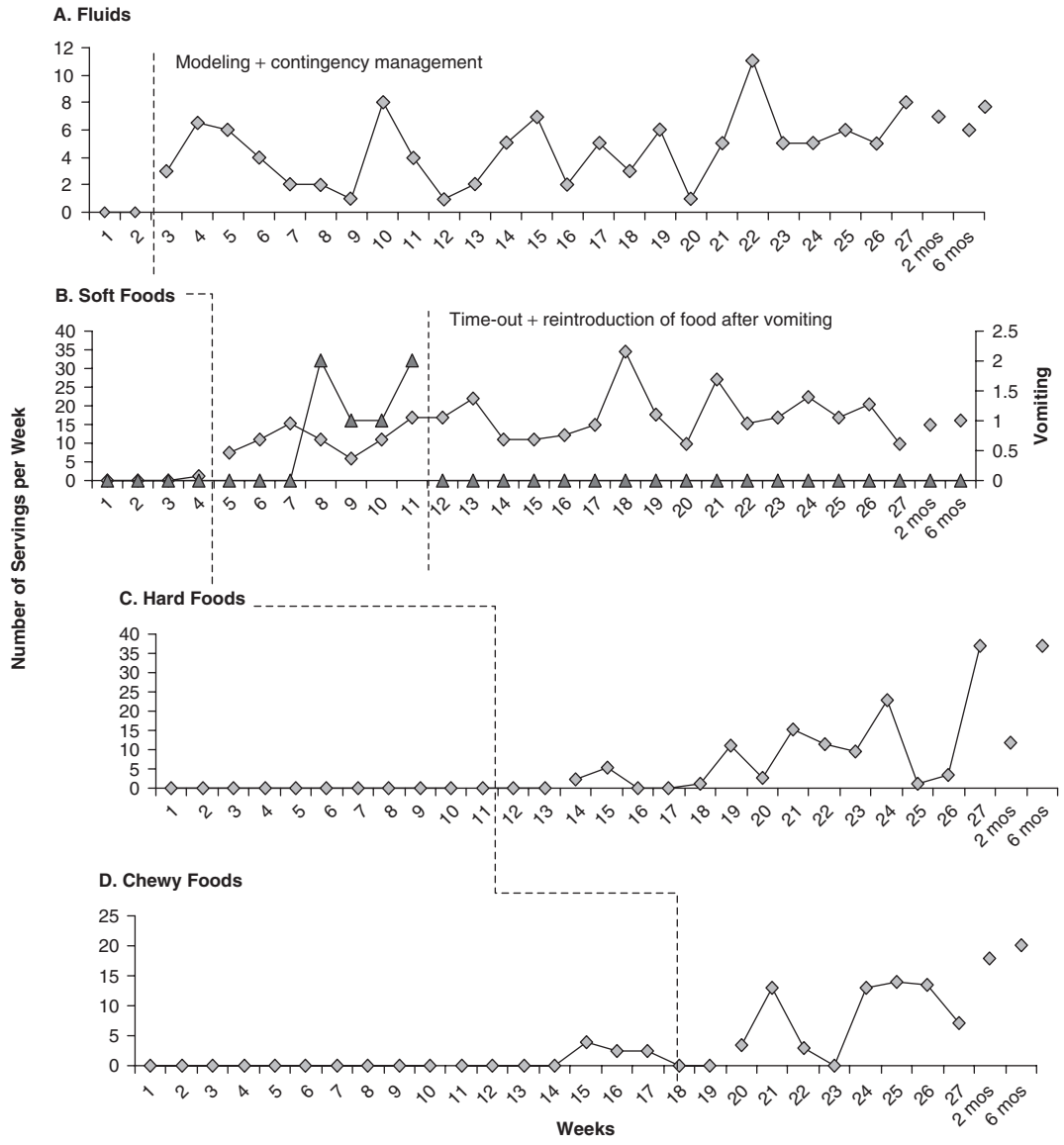

Figure 22.2 Multiple-Baseline Design Across Four Food Categories

SOURCE: From Nock, M. (2002). A multiple baseline evaluation of the treatment of food phobia in a young boy. *Journal of Behavior Therapy and Experimental Psychiatry, 33*(3–4), 217–225.

NOTE: In all four graphs, the lines marked with diamonds represent the number of servings of each food type consumed each week. The line marked with triangles in B represents the number of vomiting episodes each week (indicated on right ordinate).

Changing Criterion Designs

Changing criterion designs are those in which after a brief baseline period an intervention is directed at a behavior and the criterion for reinforcement of that behavior is changed over time to require increasing levels of behavior change. There is no withdrawal of treatment, and a causal relation between treatment and behavior change is demonstrated by showing that behavior changes when and only when the criterion for reinforcement is changed. This creates a steplike pattern of behavior change that matches the changing criterion. The decision on when to change the criterion at each step is determined by the experimenter, but typically occurs once there is

stability in the behavior. The criterion usually is changed in the same direction over time, but the experimenter may use a mini-reversal if so desired in order to increase the strength of the argument for causation. As with other designs, the intervention and the study continue until the behavior is changed and maintained at an acceptable level.

As an example, another recent study treated a 3-year-old boy presenting with food refusal, but in this study a changing-criterion design was used (Luiselli, 2000). As with the previous study, parents were taught to administer a contingency management program in the home. However, in this study the child, Sam, was rewarded with play time for consuming an increasing number of bites of food, regardless of food type. As shown in Figure 22.3, during an initial baseline period Sam consumed no bites of food. During the first intervention condition, he was rewarded whenever he consumed one or more bites of food, and his food consumption increased immediately. During the next condition, he was rewarded whenever he consumed two or more bites of food, then three, on so on. As shown in Figure 22.3, Sam's food consumption generally increased with the changing criterion over the course of the study, suggesting that the intervention caused the change in his food consumption.

Considerations. In many (perhaps most) instances, drastic changes in behavior do not occur immediately upon administration of the intervention. Changing criterion designs are a useful design option in such cases, allowing the experimenter to gradually but systematically intervene to change the desired behavior. Also, because changing criterion designs do not necessarily include a withdrawal of treatment (as in reversal designs) or an isolated focus of treatment (as in multiple-baseline designs), they require the least amount of deviation from normal clinical practices and therefore may be most easily incorporated by clinicians. A potential limitation of changing criterion designs is that it can be difficult to use them to demonstrate a causal relation between the intervention and behavior change if the target behavior does change drastically and immediately. For instance, if the criterion for reinforcement changes from zero bites of food to one bite of food in the first phase of an intervention, but food consumption changes immediately and consistently from zero to 10 bites during this

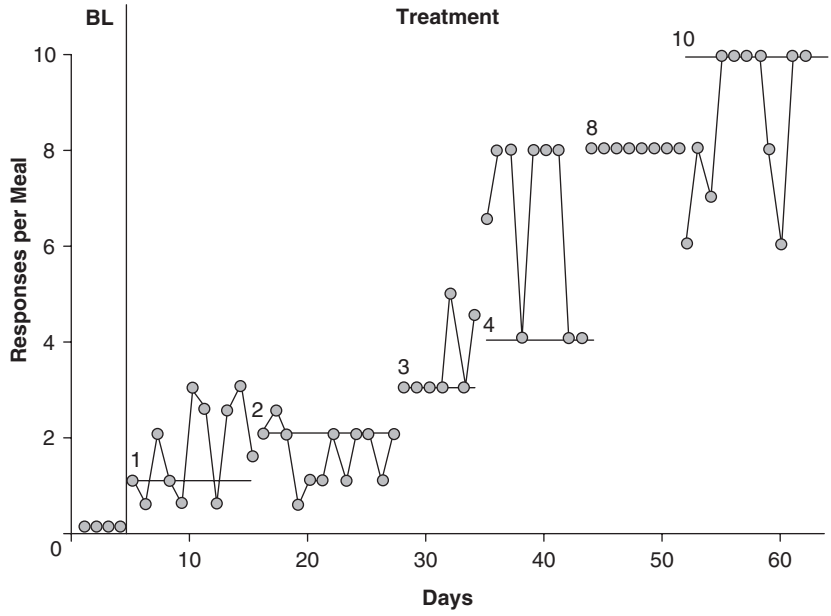

Figure 22.3 Changing Criterion Design for Treatment of Food Refusal

SOURCE: Note Luiselli, J. K. (2000). Cueing demand fading, and positive reinforcement to establish self-feeding and oral consumption in a child with chronic food refusal. *Behavior Modification, 24*(3), 348–358.

NOTE: Horizontal dashed lines preceded by numbers indicate the imposed self-feeding response criterion during meals.

phase, the argument that the intervention caused the behavior to change is weakened. Such instances would be seen as great successes clinically but limit what can be said about the reason for the change scientifically.

Data Analysis

Data collected in single-case experimental designs provide a much greater level of detail than data commonly collected in group studies. However, with this great detail comes a need for alternative methods of data analysis than those typically taught in most undergraduate and graduate methodology and statistics courses. There are two main approaches to data analysis of single-case experimental designs: those involving visual inspection of graphed data and those using inferential statistical analyses. There has been a long-standing debate about whether statistical analyses help or hinder the data evaluation process in single-case experimental designs. We will not resolve this debate here, but present a brief review of strategies from each approach along with the strengths and limitations of both perspectives.

Visual Inspection

Throughout the history of single-case experimental designs, leaders in the field have argued strongly against the need for inferential statistical tests, claiming that causal relations are not demonstrated by statistics but by careful methodological design (Skinner, 1988). From this tradition have grown methods for evaluating the effect of an intervention by graphing and visually inspecting the relation between the intervention and the dependent variable of interest.

Visual inspection of data relies on the same methodological strategy as group designs, namely, evaluating the impact of an intervention by comparing performance on the outcome of interest in a condition in which the intervention is absent with one in which the intervention is present. In group designs, this is done by summarizing the performance of many individuals in each of two or more conditions (e.g., a randomized clinical trial). In single-case designs, the data are from a single case (or small group of cases), and there are multiple observations per individual per condition, so the commonly used group design methods are not appropriate.

Evaluation of single-case data via visual inspection requires graphing the study data as displayed in Figures 22.1–3. Typically, frequency is graphed on the ordinate and time/transition through conditions on the abscissa. Kazdin (1982) has proposed four primary criteria to guide visual inspection of data from single-case experimental designs that focus on examination of behavior change across conditions. Two of these criteria relate to a change in the magnitude of the behavior: change in mean and change in level. The other two criteria relate to a change in the rate of the behavior: change in trend and latency to change.

A *change in mean* refers to the magnitude of change that occurs in the average behavioral frequency in one condition compared to another. For instance, in Figure 22.1, the average frequency across condition A_1 of 7.4 to B_1 of 1.5 represents a clear change in mean from one condition to the next. A *change in level* refers to the immediate change in behavioral performance that occurs at the transition point between one condition and the next. For instance, in Figure 1, the level of performance for the last day of A_1 is seven, and for the first day in B_1 it is two. Therefore, there is a clear and immediate change in level from one condition to the next. This change is especially clear given the behavioral stability present in each of these conditions. In single-case research designs, as in group designs, it is desirable to have minimal within-condition variability but large between-condition variability.

A *change in trend* refers to a systematic variation in the slope of the data points from one condition to the next. For instance, in Figure 22.1, there is an increasing trend in behavioral performance across condition A_2, which changes to a decreasing trend in B_2. *Latency to change* refers to the amount of time that passes between a change in condition and a change in behavioral performance. In Figure 22.1, there is no latency to change between A_1 and B_1; however, there is some latency to change between A_2 and B_2. Intervention effects are clearest when there is a stark difference in trends and no latency to change between conditions.

The visual inspection approach does not use statistical tests to evaluate each of these criteria. Instead, intervention effects should be strong and clear enough to preclude the need for statistical analyses. There is no $p < .05$ threshold and no attention to statistical significance.

Instead, only interventions with effects that are likely to be clinically significant (e.g., Jacobson & Truax, 1991; see also Chapter 21) considered effective. Interestingly, this could also be considered one of the primary weaknesses of using visual inspection to evaluate interventions. Indeed, there are instances in which small effects may be important (Prentice & Miller, 1992), and such effects would be missed if they required very strong and clear effects needed in single-case experimental designs. Another concern with visual inspection is that without the adoption of a clear and consistent threshold for demonstrating effectiveness (such as $p < .05$), there is inherent subjectivity and variability in judges' ratings of the effectiveness of study results. For example, in Figure 22.2, the transition from condition A to B to C all build support for the relation between the intervention and the observed behavior change. However, the behavior assessed in condition D increases slightly before the intervention is applied. One could argue (and one reviewer did) that this weakens support for the causal relation between intervention and behavior change. This is true, but is support weakened so much as to change the conclusion that the intervention and behavior change are causally related? The lack of a specific threshold or decision rule allows for multiple interpretations of such data. Such issues have led many to argue for the adaptation of statistical methods in the analysis of single-case experimental data.

Statistical Methods

As an alternative or supplement to using the visual inspection methods described, some single-case researchers have proposed the adaptation of inferential statistical tests for single-case research designs. The statistical approach to the analysis of single-case data historically has been used much less frequently than the visual inspection approach, but has increased in popularity and sophistication in recent years (e.g., Franklin et al., 1997).

A comprehensive review of the statistical methods used in single-case experimental designs is beyond the scope of this chapter; however, several key points warrant brief attention here. Statistical tests commonly used in group designs are not appropriate for single-case studies in cases in which observations are correlated. Correlation of data points over time is referred to as *serial dependency* or *autocorrelation* and violates the independence of error assumption of statistical tests such as the *t* and *F* tests (Matyas & Greenwood, 1997). When serial dependency exists, test values can be biased, and the researcher is advised to use statistical alternatives to these tests. Some useful alternatives include nonparametric tests such as *randomization tests,* in which treatment conditions are randomly distributed over the course of the study in order to reduce serial dependency, and the magnitude (e.g., mean) of behavioral performance for each condition can be appropriately evaluated. Another popular alternative is *time-series analysis,* in which both the magnitude and trend of behavioral performance can be compared between conditions while accounting for serial dependency (e.g., Box, Jenkins, & Reinsel, 1994).

Although the use of statistical methods to analyze data from single-case experimental designs allows for greater sensitivity to detect treatment effects, the trade-off in requiring less drastic changes is that they are more likely to capitalize on chance and to result in a Type I error. Ultimately, decisions about whether to use visual inspection of the data, statistical methods, or both should be made based on the study design and the experimenter's goals and hypotheses. Both approaches have their own strengths and weaknesses that the experimenter should take into consideration when designing and implementing the study and in interpreting the obtained data.

Strengths and Limitations of Single-Case Research Designs

Despite their limited use, single-case research designs have significant methodological advantages over large-sample research designs that could be used to facilitate more rapid clinical research advances. First, single-case research designs can demonstrate clear causal relations between intervention and behavior change with much more efficiency than large-sample designs. Indeed, large-sample studies typically require enormous amounts of time, financial resources, and staff support to implement. In contrast, any adequately trained researcher or clinician with a modest amount of resources and effort can use

single-case research designs. Given this advantage, researchers and clinicians can use single-case research designs in rapid succession to develop and evaluate individual or multiple versions of interventions, and thus can do in a matter of weeks what could take years (and millions of dollars) using large-group designs.

Second, single-case research designs offer much more flexibility in the implementation and evaluation of interventions than large-sample designs. Indeed, with a focus on maintenance of consistency of procedures across many subjects, group designs do not allow for a tailoring of the manipulation or intervention to the individual. The opportunity to modify interventions as needed provides greater research and clinical options and can lead to more innovative treatment development. Using variations on the single-case experimental designs as described previously, researchers can flexibly complete efficacy or effectiveness studies, dismantling studies, parametric studies, or any other evaluations that can be addressed using large-sample designs—and can do so within or between subjects.

Third, the assessment methods used in single-case research designs provide for the evaluation of individual change patterns in the data. Large-sample designs most often employ only pre- and post-treatment assessment, precluding evaluations of how and why individuals change over the course of treatment (see Kazdin & Nock, 2003). The use of continuous assessment and multiple experimental phases in single-case research designs allow for detailed examinations of patterns of change and the temporal relations between manipulations and their effects over time.

Fourth, although the widespread use of large-sample designs grew largely out of advances in inferential statistical methods, these very methods have come under attack due to several problems with the way they are most often used (Cohen, 1990; Krueger, 2001; Loftus, 1996; Nickerson, 2000). Most of these statistical or methodological problems are avoided or remedied through the use of single-case research designs.

Of course, single-case research designs also have clear limitations that must be considered when selecting among research design options. Many of these limitations were addressed earlier, such as the limited control on threats to validity that accompany the use of case studies and to a lesser extent quasi-experimental designs. Most of these limitations are remedied with the use of true experimental designs; nevertheless, they should be borne in mind, and efforts should be made to address them whenever possible.

The limitation most often cited in discussions of single-case research designs is a lack of generality of obtained effects. Indeed, interventions shown to be effective for a single individual may not be effective with other individuals, and these effects may not even replicate when readministered to the same individual at a later time. Although this is a clear limitation of single-case research designs, two caveats should be kept in mind. First, the use of large-sample designs does not preclude the occurrence of such problems. Indeed, what is needed for generality is the evaluation of obtained effects using different populations, conditions, or settings. Effects obtained using a homogeneous sample of individuals (which is the rule rather than the exception in large-sample studies) also may suffer from a lack of generality. Second, modifications to single-case research designs, such as the use of multiple and heterogeneous individuals within and across studies, can be implemented to demonstrate generality.

Conclusion

Single-case research designs are powerful, flexible, and efficient tools that have been largely abandoned by clinicians and clinical researchers for the past several decades. In this chapter, we summarized the methodological requirements, major design options, recent examples, challenges and limitations of case studies, quasi-experimental designs, and experimental designs used in single-case or small group studies.

There is a clear need for such designs in multiple areas of abnormal and clinical psychology. For instance, there continue to be enormous gaps between the evidence-based assessment and treatment methods developed in research laboratory settings and types of services offered to the consumer of psychological services (e.g., Weisz, Donenberg, Han, & Weiss, 1995). In addition, although efficacious treatments have been developed, questions remain about the effectiveness and mechanisms of change of such treatments (Kazdin & Nock, 2003). Work is being conducted in each of these areas; however, additional practical tools are needed to help bridge these gaps. Single-case research designs represent a natural method of doing so. These designs

provide a full menu of methods that provide the user—whether established researcher, graduate student, practicing clinician—with the power to generate and evaluate clinical hypotheses across virtually all settings, populations, and conditions. We hope this chapter encourages researchers and clinicians to consider using these methods in their daily research and practice endeavors.

References

Barlow, D. H., & Craske, M. G. (2000). *Mastery of your anxiety and panic: Client workbook for anxiety and panic (MAP 3).* San Antonio, TX: Graywind/Psychological.

Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist-practitioner: Research and accountability in clinical and educational settings.* Boston: Allyn & Bacon.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). New York: Pergamon Press.

Box, G., Jenkins, G. M., & Reinsel, G. (1994). *Time series analysis: Forecasting and control* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Cohen, J. (1990). Some things I have learned (so far). *American Psychologist, 45,* 1304–1312.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. A. Bussenius, Trans.). New York: Teachers College Press (Original work published 1855).

Franklin, R. D., Allison, D. B., & Gorman, B. S. (Eds.). (1997). *Design and analysis of single-case research.* Mahwah, NJ: Lawrence Erlbaum Associates.

Freud, S., & Breuer, J. (1895/2004). *Studies in hysteria* (N. Luckhurst & R. Bowlby, Trans.). New York: Penguin (Original work published 1895).

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19.

Kazdin, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology, 49*(2), 183–192.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*(8), 1116–1129.

Krueger, J. (2001). Null hypothesis significance testing: On the survival of a flawed method. *American Psychologist, 56*(1), 16–26.

Loftus, G. R. (1996). Psychology would be a much better science when we change the way we analyze data. *Current Directions in Psychological Science, 5,* 161–171.

Luiselli, J. K. (2000). Cueing, demand fading, and positive reinforcement to establish self-feeding and oral consumption in a child with chronic food refusal. *Behavior Modification, 24*(3), 348–358.

Matyas, T. A., & Greenwood, K. M. (1997). Serial dependency in single-case time series. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 215–243). Mahwah, NJ: Lawrence Erlbaum Associates.

Moras, K., Telfer, L. A., & Barlow, D. H. (1993). Efficacy and specific effects data on new treatments: A case study strategy with mixed anxiety-depression. *Journal of Consulting and Clinical Psychology, 61*(3), 412–420.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*(2), 241–301.

Nock, M. K. (2002). A multiple-baseline evaluation of the treatment of food phobia in a young boy. *Journal of Behavior Therapy and Experimental Psychiatry, 33*(3–4), 217–225.

Nock, M. K. (2003). Feeding disorder in infancy and early childhood. In E. Fletcher-Janzen & C. R. Reynolds (Eds.), *Diagnostic manual of childhood disorders: Clinical and special education applications* (p. 234). New York: Wiley.

Nock, M. K., Goldman, J. L., Wang, Y., & Albano, A. M. (2004). From science to practice: The flexible use of evidence-based treatments in clinical settings. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(6), 777–780.

Nock, M. K., & Kurtz, S. M. S. (2005). Direct behavioral observation in school settings: Bringing science to practice. *Cognitive and Behavioral Practice, 12,* 359–370.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112,* 160–164.

Sacks, O. (1990). *The man who mistook his wife for a hat and other clinical trials.* New York: HarperPerennial.

Schultz, D. P., & Schultz, S. E. (1992). *A history of modern psychology* (5th ed.). New York: Harcourt Brace.

Skinner, B. F. (1938). *The behavior of organisms.* New York: Appleton.

Skinner, B. F. (1988). Methods and theories in the experimental analysis of behavior. In A. C. Catania & S. Harnad (Eds.), *The selection of behavior: The operant behaviorism of B. F. Skinner: Comments and consequences* (pp. 77–105). New York: Cambridge University Press.

Watson, J. B. (1925). *Behaviorism.* New York: Norton.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*(5), 688–701.