

## Chapter 13

# ***APPLICATION OF TIME-SERIES (SINGLE-SUBJECT) DESIGNS IN CLINICAL PSYCHOLOGY***

SCOTT T. GAYNOR, M.A., SUSAN C. BAIRD, M.A., and ROSEMERY O. NELSON-GRAY, PH.D.

A gulf exists between research and practice in clinical psychology (Rice, 1997). It has been suggested that researchers are not sufficiently guided by clinical discoveries, and that practitioners, though valuing research findings, do not access these from traditional research journals (Beutler, Williams, Wakefield, & Entwistle, 1995). Reflecting on the centennial of clinical psychology, Meehl (1997) commented on the fallibility of anecdotal evidence (i.e., clinical experience) and the need for quantitative research. Practitioners, however, often struggle to generalize from large-scale, quantitative research findings, where results are reported as probabilities of efficacy across groups, to the unique situation of an individual client (Stricker & Trierweiler, 1995). Thus, Stricker and Trierweiler call for a local clinical scientist model, whereby practitioners apply not only general research findings, but also employ a scientific attitude, or scientific thinking, in individualizing treatment for specific clients in the local setting.

The issues introduced above could be addressed if a methodology existed that is pragmatic and useful for the practicing clinician, more scientifically rigorous than anecdotal impressions, and sensitive to the individual nature of clinical cases. We argue that time-series (i.e., single-subject) designs are such a methodology.<sup>1</sup> This chapter explores the major reasons to use these designs, the fundamental information needed for their application, and the different component elements of which they are comprised. Where possible, examples from the clinical literature are employed. This chapter is written to highlight the basic knowledge needed to implement single-subject designs and to provide numerous examples of their use. As such, space does not allow for an exhaustive account of all the methodological subtleties involved in single-subject designs. The interested reader is directed to books by Barlow and Hersen (1984), Hayes, Barlow, and Nelson-Gray (in press), and Kratochwill (1978) for more exhaustive analyses.

---

<sup>1</sup> Time-series and single-subject will be used interchangeably in this chapter. The former is somewhat preferable as it places the focus on the design rather than the number of subjects involved, as time-series designs often involve more than one subject.

## WHY USE TIME-SERIES METHODOLOGY?

The reasons to employ time-series methods include both the practical and the empirical. Practically, demonstrating accountability and cost-effectiveness of treatment is becoming increasingly necessary in the current managed care environment (Giles, 1991). Time-series methods can enhance clinicians' ability to demonstrate the effectiveness of their interventions without a great deal of additional effort. Practitioners, in their clinical decision making, likely use a similar rationale to that recommended in time-series designs (Hayes, 1981). For instance, initial sessions with the client are likely to focus on assessment (i.e., determining the frequency, intensity, duration, and historical context of the client's difficulties). This is the baseline against which the intervention will be judged. The outcome of this initial assessment is a case conceptualization, sometimes called a functional analysis, which points (implicitly or explicitly) toward the critical dependent variables that will be the focus of treatment. The subsequent treatment is the independent variable. Broadly speaking, time-series designs simply involve more systematic delineation of these aspects of clinical intervention.

As well as being consistent with the rationale of routine clinical practice, time-series designs may actually enhance clinical practice. The systematic nature of time-series designs may increase precision in making assessment and treatment decisions. For instance, frequent objective measures provide feedback for the clinician that allows for the potential alteration of ineffective interventions and the continuation of effective ones.

Empirically, time-series designs can address many of the scientifically important questions raised in the clinical environment (e.g., Does a treatment work? Which of two treatments is most effective? Are they both effective? What are the "active" components of a treatment?) (Hayes, Barlow, & Nelson-Gray, in press). In addition, time-series designs seem to provide a direct mechanism for the integration of research and practice. On the one hand, researchers often use large numbers of homogeneous subjects to demonstrate the efficacy of interventions. The generality of these efficacious treatments, however, is often less well established (Seligman, 1995). This is where practicing clinicians could make important contributions, using more idiographic and flexible, yet rigorous, time-series designs (Hayes et al., in press). On the other hand, clinicians, especially those of a psychodynamic orientation, find that "the traditional case report remains our most compelling means of communicating clinical findings" (Spence, 1993, p. 37). Logically, however, to go from a traditional case study to a more scientifically rigorous time-series design simply involves increased quantification of the dependent variable and greater specification of the independent variable, which is increasingly becoming recognized by psychodynamically oriented researchers and clinicians (e.g., Fonagy & Moran, 1993). Finally, the implementation of time-series designs may help increase the likelihood that researchers are informed and influenced by the expertise of practicing clinicians.

## FUNDAMENTALS OF TIME-SERIES METHODOLOGY

The objective when using time-series methodology, like any experimental methodology, is to distinguish the effects that result from a given intervention (i.e., the independent

variable) from effects that may be caused by unrelated variables (extraneous variability or error in measurement). In short, the objective is to rule out threats to internal validity. When the effects can clearly be attributed to the independent variable, the experiment is internally valid (Kazdin, 1980). In this broad sense, time-series methodology is no different from group comparison approaches. The major distinction between time-series and group designs is that in the former, the effects are analyzed at the level of the individual. This requires that data be collected in a fashion that facilitates the making of valid inferences at the idiographic level (Kazdin, 1981). In this section, we highlight some of the major issues and design characteristics influencing the likelihood that valid inferences can be drawn. These are also summarized in Table 13.1.

### Type of Dependent Measures

Time-series designs require the use of objective dependent measures that accurately and sensitively measure the important units of behavior targeted for change (e.g., as determined in the initial assessment and conceptualization of a clinical case). The clinically important units may include client actions, or overt behavior; cognition, or verbal behavior; and/or physiological responses (Nelson, 1981). For instance, Ferguson and Rodway (1994), in a study of cognitive-behavioral treatment for perfectionism, measured both perfectionistic thoughts (via two questionnaires) and instances of perfectionistic behaviors (agreed upon by the therapist and the client).

The collecting of dependent measures should begin early in therapy, possibly before the first session with the client. For instance, Beck, Rush, Shaw, and Emery (1979) mailed initial assessment measures to clients to bring completed to the first session. In many clinical settings, clients spend a period of time on a waiting list, time that could

**Table 13.1. Characteristics of the design and major threats to internal validity to be considered in drawing inferences from time-series data**

---

*Characteristics of the Design*

Objective data  
 Continuous assessment  
 Stability of problem  
 Immediate and marked effects  
 Replication with multiple subjects or reversals

*Major Threats to Internal Validity*

Coincidental external/extraneous events  
 Maturation/learning (gradual biological or psychological processes occurring within a person over time)  
 Testing/assessment (reactivity; potential changes as the result of assessment)  
 Statistical regression (regression to the mean)  
 Variability  
 Multiple intervention interference (order effects, carryover effects, alternation effects)

---

*Source:* Adapted from Kazdin, 1981; Hayes et al., in press.

be used for assessment. Taking measures early allows for an adequate baseline to be established while not unnecessarily delaying treatment.

In addition to beginning data collection early, multiple measures that are both global and specific should be taken initially. The use of a broad range of measures early in the course of therapy is consistent with the general clinical practice of using the first sessions for initial assessment and information gathering. Taking multiple measures may appear cumbersome as some measures are lengthy and time consuming. However, there are several remedies for this. First, after the initial assessment, some areas that are not targeted for treatment will no longer need to be assessed. Second, more practical measures can be utilized more frequently and global, lengthier, or more difficult measures less often. In a study combining cognitive therapy and interpersonal therapy for depression, Jensen (1994) had subjects complete the 21-item Beck Depression Inventory (BDI; Beck et al., 1979) weekly, whereas the more global Social Adjustment Scale (SAS; Weissman & Paykel, 1974) was completed only at pre- and posttreatment.

The Jensen (1994) study highlights another critical point: Measures need to be taken repeatedly over the course of the initial assessment, during treatment, and, ideally, at follow-up after treatment. The quality of the measure as well as its practical utility should be considered in determining which measures to use most often (see Fischer & Corcoran, 1994, for actual questionnaire measures covering a wide range of clinical symptoms and for use with adults, children, couples, and families; see also Nelson, 1981, for discussion of a broad range of realistic dependent measures involving self-monitoring, direct observation, self-ratings, and self-report).

The final point regarding the taking of dependent measures relates to the situation specificity of behavior (Mischel, 1968). Because a client's responding may be situationally determined, the conditions under which dependent measures are taken should be similar across measurement instances to the extent possible. For instance, Ferguson and Rodway (1994) asked clients to complete each weekly measurement package in the same location and at the same time during the day.

There are several cautions in order when taking repeated measures with the same instrument. Many of these cautions apply any time dependent measures are taken and are not unique to repeated measures. In general, the concerns described below involve alternative explanations for observed changes and are thus threats to internal validity (see Table 13.1). One potential problem involves reactivity to the measurement process. Reactivity is defined as behavioral change that results simply from the awareness that one's behavior is being monitored. As this change in behavior is not the result of the intervention, but may appear as such, it is a threat to internal validity. However, as reactivity to measurement generally occurs at the onset of measurement, the collection of a careful baseline provides a protection against attributing such effects to the intervention. It is also important to reduce demand characteristics that may be associated with assessment. That is, clients may report positive therapeutic effects when in fact such results are absent. For instance, after several weeks of treatment, a client may feel obligated to report some positive changes for reasons such as feeling warmly toward the therapist, wanting to please the therapist, or because he or she is paying for the service rather than because improvement has actually occurred. Taking multiple measures of different types (e.g., global and specific as well as self-report and direct observation) and discussing the measurement process, including the client's reaction to and experience of it, may reduce the

likelihood that demand characteristics significantly impact the results. A final concern results from the tendency for extreme scores to revert toward the mean without intervention. Such an effect is called regression to the mean. For instance, clients presenting to therapy in extreme distress may report a reduction in symptoms over the course of treatment that may be better accounted for by a return to previous levels of functioning as opposed to effective treatment. A constant-series control (see below) may provide some protection against statistical regression. The use of a constant-series control allows for a between-series comparison of treatment versus no-treatment, controlling for regression to the mean, which should have a comparable impact in both series. Also, establishing a relatively stable preintervention baseline provides some protection against regression to the mean. That is, the more similar the baseline scores within a series, the more unlikely it is that they are simply transient extreme scores, which would subsequently regress to the mean. It is to the topic of establishing a preintervention baseline that we now turn.

### **Establishing a Preintervention Baseline or A Phase**

Measurement typically begins prior to intervention, and this is called the baseline or A phase (Barlow & Hersen, 1984). The baseline provides the basis with which the intervention (B phase) will be evaluated. That is, the baseline involves repeated measures of the client's behavior as it is maintained in the absence of intervention. The rationale is that without intervention the baseline should continue and, therefore, provide a prediction of what would occur in the absence of treatment.

An ideal baseline allows for the assessment of level, trend, and variability (Hayes et al., in press). Level is essentially some measure of the magnitude of the dependent variable, such as a BDI score of 30 (BDI scores between 0 and 9 are considered normal, 10 to 20 mild depression, 20 to 30 moderate depression, and greater than 30 severe depression [Kendall, Hollon, Beck, Hammen, & Ingram, 1987]). Trend refers to the pattern of the dependent variable during baseline and is analogous to slope. For instance, BDI scores of 20 and 40 over two consecutive weeks of baseline would suggest a trend toward increasing symptoms of depression (a steep slope). However, BDI scores of 32, 27, and 31 suggest very little trend in the data (a shallow slope). Variability in the dependent measures collected influences our confidence in the estimates of the level and trend of the behavior. For instance, a client with baseline BDI scores of 32, 27, and 31 provides a more stable baseline than a client with BDI scores of 20, 40, and 30.

The minimum number of data points needed for an assessment of level, trend, and variability is three (although level can be assessed with one data point and trend with two) (Hayes et al., in press). The more data the better, as this provides a comprehensive picture of what is occurring in the absence of treatment and is used as a predictor of future client functioning with which treatment results will be compared. The clinical environment often limits the amount of baseline data that can be gathered. That is, in some clinical situations, such as with a client expressing suicidal ideation, immediate treatment is required. At a minimum, the level, trend, and variability in baseline must allow for an effect to be seen if it occurs. Assessment of baseline functioning can serve clinically and scientifically useful purposes. If the level of a client's responding to the BDI, for example, is repeatedly in the normal range, an intervention targeting depression may not be warranted practically and effects may not be visible. Similarly, if the trend in the

baseline data is in the direction expected by the intervention, it makes it more difficult to determine treatment effects. The top panel of Figure 13.1, adapted from Ferguson and Rodway's (1994) study of the treatment of perfectionism with cognitive-behavioral therapy, illustrates such a case. Notice how the major decrease in irrational values occurs prior to the initiation of treatment. A substantial additional decrease would be needed to demonstrate a treatment effect. One is also left wondering what was occurring during the first two assessment periods for this client. It is possible the individual had several particularly difficult days (possibly due to some coincidental events) that temporarily elevated the scores, which subsequently decreased to reflect the individual's more average functioning (regression to the mean) throughout the remainder of baseline.

The middle panel of Figure 13.1 demonstrates a highly variable baseline that stabilizes when treatment begins and decreases in Week 7. Confidence in the treatment effect is bolstered because the latter data points fall clearly outside the range of the variable baseline. However, the potential effect of extraneous events occurring in Week 7 needs to be ruled out as an explanation for the abrupt decrease at that time. Ideally, the abrupt decrease would occur earlier in treatment. This point emphasizes the tendency to place more confidence in effects that are large in magnitude and occur immediately when the intervention is implemented. The effects pursued in clinical work are often more gradual, however. In addition, some treatments are presented over several sessions, and, therefore, an immediate effect would not be expected. In these cases, other strategies (e.g., replication) are available to bolster confidence that the intervention produced the gradual or delayed effects.

There is no one solution to dealing with excessively variable baseline data. Hayes et al. (in press) make four recommendations: analyze potential sources of variation, continue baseline, evaluate the level/unit of analysis being used, and proceed with the intervention anyway. Notice, in the middle portion of Figure 13.1, that Ferguson and Rodway (1994) decided to proceed with the intervention despite the variability. This decision may have been made because they thought that even with the variability an acceptable treatment effect could be demonstrated if it occurred, and/or they felt an ethical responsibility to begin treatment. Interestingly, Ferguson and Rodway took baseline data every three days and intervention data weekly, thus using somewhat different units across the conditions. Blocking (or "chunking") data in baseline so that the unit is now a weekly composite reduces the variability significantly and reveals a moderate downward trend (see bottom of Figure 13.1). Altering the unit of analysis is not a trick to turn poor data into good data. A rationale should always be provided. In this case, the rationale was that different units were being compared across conditions and that the additional measures taken in baseline may have increased the likelihood that extraneous daily factors were included in these data that would be less likely to be included in a weekly measure.

### Implementing the Intervention or B Phase

As discussed previously, the goal of time-series designs is to rule out threats to internal validity, thereby strengthening confidence that any effects noted are the product of the intervention (i.e., independent variable). Thus, it is important that only one independent variable be introduced at a time (Hayes, 1981). If two or more independent variables are

