Journal of Consulting and Clinical Psychology
1981, Vol. 49, No. 2, 193-211

0022-006X/81/4902-0193$00.75

Single Case Experimental Design and Empirical Clinical Practice

Steven C. Hayes
University of North Carolina at Greensboro

Research in clinical psychology is done very infrequently by the practicing clinician. A major reason for this seems to be inadequate or cumbersome research tools that are incompatible with clinical realities and assumptions. Time series experimentation is explored as a possible research tool available to clinical practice. Standards of good clinical decision making seem to parallel closely the logic of time series methodology. It is argued that most of the reasons for the underutilization of this methodology in clinical practice have to do with misunderstanding and biases on the parts of clinicians and methodologists alike. Time series experimentation is broken down into several logical core elements and organized into an overall system according to the nature of the predictions against which comparisons are made. The natural use of these logical steps in clinical practice is examined in terms of its practical, scientific, and ethical dimensions.

The progress of research in clinical psychology presents something of a paradox. The social need for clinical research can hardly be overestimated; the field incorporates many of the most serious social and personal ills of the day. Further, tremendous resources are available to the field in the numbers of professionals, training programs, employment opportunities, and (compared with many disciplines) funding patterns. Yet data abound that these needs and resources have not yet been fully combined to produce maximum research progress (Garfield & Kurtz, 1976; Kelly, Goldberg, Fiske, & Kilkowski, 1978; Levy, 1962).

This paradox has often been noted, especially in the well-worn discussion of the research/practice split (e.g., Leitenberg, 1974; Meehl, 1971; Peterson, 1976; Raush, 1969, 1974; Rogers, 1973; Shakow, 1976). Some psychologists have rationalized the split, pointing to the irrelevance of the traditional research enterprise to clinical practice (Holt, 1971; Meehl, 1971; Peterson, 1976; Raush, 1974). Others have denied the split, defending the scientist-practitioner model (e.g., Shakow, 1976) and calling for better, more controlled, and even more intricate clinical research (e.g., Meltzoff & Kornreich, 1970; Paul, 1969; cf., Thelen & Ewing, 1970). A third reaction has begun to receive some attention (e.g., Barlow, 1980). It attempts to dissolve the split, claiming that practicing clinicians may not be lacking a dedication to research, just tools for the task. If single case (or time series) methodology[1] could be taught in a manner

I would like to thank Rosemery O. Nelson for her helpful comments.

Requests for reprints should be sent to Steven C. Hayes, Department of Psychology, University of North Carolina, Greensboro, North Carolina 27412.

[1] The terminological diversity surrounding this research strategy is enormous. These designs have been termed single subject, $N = 1$, or single case (e.g., Hersen & Barlow, 1976); intrasubject replication (Kazdin, 1980); intensive (Chassan, 1967, 1979); own control (e.g., Millon & Diesenhaus, 1972); and time series (e.g., Campbell & Stanley, 1963; Glass, Wilson, & Gottman, 1975), among other names (Jayaratne & Levy, 1979). I have chosen to use two terms somewhat interchangeably. The first is *time series experimentation*. It emphasizes the critical component of these designs. Its drawback is possible confusion with time series analysis, a statistical technique used to analyze time series data (e.g., Box & Jenkins, 1976; Gottman, McFall, & Barnett, 1969), or confusion with specific designs, such as Campbell and Stanley's name for an A/B design. The other term, *single case designs*, emphasizes the number of subjects as the central issue. For clinical work, analyzing the individual is a desirable end in and of itself (Bernard, 1865), and this is the most popular name for these designs. Nevertheless, many of these designs (e.g., multiple baseline across subjects) require several subjects, and all can be done with entire groups as the unit of analysis. Other terms are more problematic. "Intensive" carries an evaluative connotation. "Intrasubject replication" and "own control" wrongly assert that all

that fits demands of the clinical environment (the thinking goes), practicing clinicians could produce more research data and make consumption of clinical research more worthwhile for the practitioner.

This view has been advanced periodically over the years (e.g., Barlow & Hersen, 1973; Browning & Stover, 1971; Chassan, 1967, 1979; Hersen & Barlow, 1976; Kazdin, 1978, 1980; Leitenberg, 1973; Svenson & Chassan, 1967). Most of the conceptual work to date, however, has been oriented toward the full-time clinical researcher, not the practicing clinician. Clinical researchers and academic clinicians have not been unresponsive to single case methodology, but group comparison approaches are often equally attractive and valuable. It is in the on-line clinical environment that the unique value of time series experimentation truly becomes apparent, yet little has been done to advance its use there.

The goodness of fit between clinical decision making and time series methodology is remarkable. As will be shown, good clinical practice seems often to be a type of single subject experimentation in that the logic of the two enterprises is so similar. The present article will argue that good practicing clinicians are already doing evaluations of potential scientific value with most clients they see. They need only (a) take systematic repeated measurements, (b) specify their own treatments, (c) recognize the design strategies they are already using, and (d) at times use existing design elements deliberately to improve clinical decision making.

If this argument can be shown to be correct, then it is worth considering why single case experimentation, hardly a newcomer on the methodological scene, is so underutilized in applied settings. Several reasons might be suggested:

1. It is undertaught. In most training programs, methodological courses are taught by nonclinicians (e.g., statisticians, general experimental psychologists). With some notable exceptions (e.g., the experimental analysis of behavior), most of these other subfields are heavily committed to group comparison research.

2. It has not been aimed at the practicing clinician. Perhaps in order to show that time series methodology can be just as scientific as group comparison approaches, methodological niceties have often been overemphasized. Individual clinicians cannot be expected to distinguish between the core essentials and simple issues of degree, and it may be rejected because it is seemingly impractical to do it right.

3. It is associated with behaviorism. Historically, single case methodology has been most heavily developed and used by behaviorists (e.g., Sidman, 1960) and may often be rejected because of it. This is unfortunate, however, because the methodology is theory free. One can use time series experimentation to answer questions about self-disclosure as readily as behavioral indicants of anxiety, and about insight-oriented procedures as successfully as assertiveness training.

4. Clinicians may fail to distinguish between research methodology and group comparison approaches. To most clinicians, group comparison research is research. Individual clinicians (and clinical training programs) are likely to throw the single case baby out with the group comparison bathwater.

5. There are few outlets for on-line clinical research. On-line single case evaluations, modified as they frequently are by realities of clinical practice, may meet a severe reception in most clinical journals. Reviewers of such articles are themselves unlikely to be practicing clinicians, and appropriate standards for evaluations of actual clinical practice are still unformed.

6. Clinical agencies often provide little support for scientific work. Everything from case loads to secretarial help to agency policies concerning research may hinder on-line use of single case methodology. Fortunately, third-party payments are beginning to create counter pressures for clinical evaluation.

In the past few years, a whole host of professional developments have indicated the possible beginning of an empirical clin-

control strategies in these designs are within subject when many of them (e.g., baseline-only control, multiple baseline across subjects) are not.

ical movement based on the combination of single case methodology with the resources of the practicing clinician. These include books (e.g., Jayaratne & Levy, 1979), articles (Barlow, 1980; Levy & Olson, 1979; in fact, the present series of articles), conferences (e.g., the Association for Advancement of Behavior Therapy adopted this issue as the theme of its 1980 convention), special issues of journals (e.g., a 1979 issue of the *Journal of Social Service Research*; an upcoming issue of *Behavioral Assessment*), workshops, and the like.

The present article will outline the nature of time series experimentation and underline ways it can fit into routine clinical practice. I have assumed that the clinician is working with actual paying clients (individuals, groups, agencies, etc.) who have given consent to routine clinical evaluation and treatment. No willingness to endure a significant increase in risk, cost, or time to the client beyond that required by good clinical decision making is assumed. Finally, I have not attempted to analyze in detail the many fine points and issues raised by single case experimentation (for that, the interested reader is referred elsewhere, e.g., Hersen & Barlow, 1976; Jayaratne & Levy, 1979; Kratochwill, 1978; Sidman, 1960), and general, noncontroversial recommendations drawn from such sources have not been referenced separately.

The Essentials of Single Case Methodology

All time series work is based on combining essential core elements into logical designs. In this section, the general rules of approach will be described. In the following sections, specific design elements will be detailed.

Repeated Measurement

The absolute core of time series methodology, as denoted by the very name, is repeated measurement of the client's behavior, including thoughts, feelings, and so forth. Because estimates of the stability, level, and trend of the data (against which treatment effects might be seen) are drawn within subject, the clinician must have a record of client progress across time (see Nelson, 1981). Repeated measurement also parallels rules of clinical practice. Practical clinical guides often exhort clinicians to "examine regularly and consistently whether therapy is being helpful" (Zaro, Barach, Nedelmann, & Dreiblatt, 1977, p. 157).

In clinical practice, repeated measurement should start early, using several measures if possible. An experienced clinician often has a good idea of several of the client's problems even before the end of the first session. If measurement is begun immediately, then when normal assessment ends, the clinician will often have a systematically collected baseline. Early collection of systematic measures will also often contribute to clinical assessment itself. Some problems, when measured repeatedly, will turn out not to be real difficulties. Measures should also be practical. It is better to collect measures of medium quality than to collect none because excessively high standards of measurement are set. Finally, they should be taken under reasonably consistent conditions to avoid variability caused by inconsistent measurement procedures.

Establishment of the Degree of Intraclient Variability

An estimate of the degree of variability in the client's behavior (as repeatedly measured) is critical in single case methodology. In the context of this estimate, determinations are made about the level and trend in the behavior, and predictions are drawn about the future course of the behavior. Measures need only be stable enough to see effects, should they occur. The target problem and probable effects of intervention bear heavily on issues of stability, the question is always, Stable in terms of what? For example, if a total reduction of a behavior is anticipated, extreme variability would present no problem. Conversely, if measurement variability could not allow any treatment effect to be seen, then it would be foolish to proceed. This methodological advice dovetails nicely with clinical realities, however. For example, a client showing infrequent nondestructive outbursts of anger

would probably not be treated for anger control if their frequency would be indistinguishable from that expected after treatment.

When the client's behavior is excessively variable, several actions can be taken. First, the clinician can simply wait until patterns become clearer. Often variability is temporary (for example, it may be caused by the initial effect of entering treatment), and it is frequently better to wait than to plunge ahead unnecessarily.

Second, if at times the client is behaving well and at times badly, the practicing clinician will probably begin to search for factors that account for these differences. For example, if a client's self-esteem (as measured, say, by a brief paper-and-pencil instrument before each therapy session) is high some weeks and very low others, the clinician may search for reasons accounting for it. Finding that the client's self-esteem is low only on weeks following rejection by potential dates might lead to a treatment program of social skills training or therapy around the issue of rejection. Further, the previously unstable measures might now be quite stable when organized into times following or not following instances of rejection.

A third strategy is to examine the temporal unit of analysis. Often measures are collected in particular units for convenience (e.g., clients are often asked to self-record in daily blocks). If the actual phenomena may be better seen in larger units, then the data may be blocked (or intraclient averaged). For example, a clinician working with a marital couple might find that daily records of arguments reveal extremely variable behavior, some days there are no arguments, and on others there are several. This may be expected, since all couples have some good days and some bad. More clinically important may be, for example, the average number of arguments in a week. When the data are blocked by weeks, stability may emerge. Some of the detail is lost, but this is always true: Organizing events by day disguises hourly variability; organizing them by hour disguises minute-by-minute effects. Part of good clinical skill seems to involve knowing when to ignore individual trees in order to see the forest.

A final strategy is to proceed anyway. If the effects are very strong, they may be seen. If not, enough may be learned that the next client may benefit.

Specification of Conditions

All research requires clear specification of the independent variable. In the clinic, true "technological" (Baer, Wolf, & Risley, 1968) specificity is sometimes difficult. Even when the therapist cannot specify the intervention, however, it may be possible to measure therapist behavior using some of the same within-clinic procedures for measuring client behavior (Nelson, 1981; for example, see Becker & Rosenfeld, 1976).

Replication

The logic of all time series designs requires replication of effects. In the clinic, this requirement is increased because of the methodological compromises often forced there. In addition, the external validity of single case research depends on systematic replications of effects in many clients.

An Attitude of Investigative Play

Undoubtedly the biggest difference between group comparison research and time series methodology is the overall approach that they encourage. Single case research should be a dynamic, interactive enterprise in which the design is always tentative, always ready to change if significant questions arise in the process. The data should be graphed frequently and in various forms so that apparent patterns can emerge and leads can be followed. Group comparison research, however, is usually planned out in detail beforehand and then simply carried out.

One of the common mistakes made by researchers in time series research is their approaching these tools as they would approach group comparison research (e.g., deciding beforehand on a sequence of phases or setting specific phase length). Unfortunately, clients' data often do not conform to the preset mold; these data often do not confirm preset hypotheses. When unanticipated effects are seen, the clinician must be ready to abandon previous design decisions and to

let the client's data be the guide. This is also good clinical practice. For example, clinical guides advise clinicians to "be prepared to alter your style of dealing with a client in response to new information" and "be prepared to have many of your hypotheses disproved" (Zaro et al., 1977, p. 28).

Other Suggested Requirements

Many other rules about single case methodology are not essential but are issues of degree. One rule is to keep phases about the same length (Hersen & Barlow, 1976). Widely different phase lengths can produce errors in interpretation, but changing phases based on time alone can also produce unclear comparisons. This is a matter of degree, and its importance can be minimized by clear effects and systematic replication. Another rule is, "Change one variable at a time" (Hersen & Barlow, 1976). This rule is often a good one, but it can be easily misinterpreted (e.g., Thomas, 1978). The meaning of "variable" here is better conveyed by the phrase "condition you wish to analyze." Thus, entire packages may be varied when it is the package that is being evaluated.

Creative Use of the Design Elements

The creative use of time series designs may have been inadvertently hindered by the literature's emphasis on complete designs rather than on design elements. For example, designs such as an A/B/A or B/C/B/C have often been described as separate designs even though their logical structures are identical (e.g., Hersen & Barlow, 1976; Mahoney, 1978; indeed, virtually the entire literature in the area has followed this course). All single case designs are built from a small set of building blocks. There are potentially as many specific single case designs as there are designs for brick buildings, and the core elements of each are comparably simple.

The present article distills all time series work into a few core elements, organized by the nature of their estimates of stability and the logic of their data comparisons. These core elements can then be creatively combined to contribute to good clinical decision making. There are three general types of strategies used: within, between, and combined series. All current single case design elements can be readily organized into these three types.

Within-Series Strategies

The best known types of time series elements rely on changes seen within a series of data points (in a single measure or homogeneous set of measures). There are two subtypes of within-series strategies: the simple phase change and the complex phase change. Each of these will be described, and their use in clinical practice will be detailed.

The Simple Phase Change

The cornerstone of many of the most popular single case designs is the simple phase change. This element consists of (a) the establishment of stability, level, and trend within a series of data points across time, taken under similar conditions; (b) a change in the conditions impinging on the client; and (c) examination of concomitant changes in the stability, level, or trend in a series of data points taken under the new conditions. It is a within-series strategy in the sense that it is systematic changes seen within a series of data points across time that are examined.

A common example of the simple phase change is the A/B design. If the stability, level, or trend shown in A suddenly changes when B is implemented, our confidence increases that B is responsible for that change. Often there are possible alternative explanations for the effect (e.g., maturation, the effect of measurement, coincidental external events; see Campbell & Stanley, 1963; Hersen & Barlow, 1976; Kratochwill, 1978), and usually the effect must be replicated before our confidence in the effect is sufficiently high. One way is to repeat the phase change in reverse order (the A/B/A design). If the behavior tracks the change once again, our confidence increases further. This simple phase change process can be repeated indefinitely, each sequence forming a new completed design (e.g., A/B/A/B; B/A/B). Two treatments can be compared in the same manner (e.g., B/C/B; C/B/C/B). All

of these are merely specific applications of the logic of the simple phase change, allowing us to ask questions such as, Does treatment work? or Which treatment is better?

Complex Phase Changes

The simple phase change can be coordinated into a more complex series of phases. Each of the complex phase change strategies specifies an overall integrative logic.

Interaction element. This is a series of phase changes in which a treatment or treatment component (B) is alternately added or subtracted from another treatment or treatment component (C). A number of specific sequences are possible (e.g., B / B+C / B; C / C+B / C; B+C / C / B+C). Its logic is essentially identical to the simple phase change. This can be easily seen if instead of writing A/B/A one were to write the equally correct A / A+B / A. The question, however, seems a bit more complex, namely, What is the combined effect of two treatment components compared to one alone? As an example, suppose a clinician wonders if the empty-chair technique is really helpful in the treatment of unresolved grief. In the first phase, a specified set of techniques (B+C) might be used, including empty-chair exercises involving the lost loved one. This technique (C) might then be withdrawn and reinstituted, forming a B+C / B / B+C design. If the client's functioning tracks these changes, the role of this procedure in the overall package could be determined.

Combining Does B work? and Does C work? elements. A simple phase change comparing two treatments does not make sense unless it is known that either works relative to baseline. If this is not known, the design must compare them with baseline as well as with each other by combining simple phase change strategies for determining their effectiveness. For example, the sequence A/B/A/C/A combines an A/B/A with an A/C/A. This allows us to ask if B and if C are effective. It also allows a comparison of the two treatments, but it is weak, because order effects are possible and noncontiguous data are being compared (the data in the B phase with those in the C phase). To strengthen this comparison, other subjects might receive an A/C/A/B/A sequence. If the conclusions are the same, then the believability of the treatment comparison is strengthened.

Changing criterion. This element (see Hartmann & Hall, 1976) is usually based on the following line of reasoning: If you arbitrarily specify the level that a given behavior must reach to achieve an outcome, and the behavior repeatedly and closely tracks these criteria, then the criteria are probably responsible. Typically, this element is used when the behavior can only change in one direction, either for ethical or practical reasons. The logic of the maneuver, however, allows for criterion reversals when the behavior is reversable.

The weakness of the procedure is that it is not always clear when observed behavior is tracking criterion shifts. This problem can be alleviated by altering the length and magnitude of criterion shifts (or, if possible, their direction), as shown in Figure 1.

Other strategies. Several other complex phase change strategies exist, although they are used infrequently in the applied literature. For example, an ascending/descending design element (see Sidman, 1960) is a popular research tool in basic operant psychology.

Using Within-Series Strategies

When a clinician begins to work with a client, be it an individual, group, or agency,

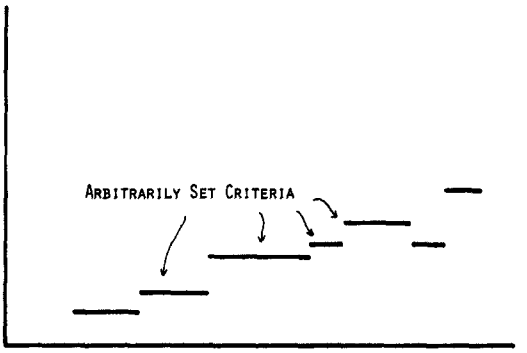


Figure 1. An example of the arbitrary manipulation of the length, depth, and direction of criterion shifts, making any behavioral correspondence with the criteria more obvious.

it is rare that an elaborate clinical question springs forth in whole cloth. Clinical work usually involves a gradual process of investigation. The use of within-series strategies provides a good example of how single case methodology suits itself to this clinical reality. In the sections that follow, the sequence of events faced by a clinician doing a within-series evaluation will be described. The choice points and design options in this process will be given particular emphasis.

Establishing the first phase. The clinician typically begins a therapeutic relationship with a period of assessment. If the advice offered earlier has been followed, when this period ends, a baseline is already in hand or nearly so. Several rules have been offered as to the adequacy of obtained baselines.

A first consideration is the length of the first phase. To establish estimates of stability, level, and trend, at least three measurement points seem to be needed (e.g., see Hersen & Barlow, 1976), though more are desirable. If fewer have been obtained, and the needs of the client are clear, then the practicing clinician must push ahead anyway. To do otherwise would be to delay treatment for research, not clinical, reasons. Short baselines are not necessarily lethal. There may be other information available about the problem being measured. For example, the disorder may have a known history and course (e.g., the social withdrawal of a chronic schizophrenic), or archival baselines may be available (e.g., records from previous therapists). Also, the clinician can often make up for short baselines by using other design elements later (e.g., withdrawals) or by replicating the effects with others (e.g., multiple baselines across subjects).

A second consideration is the stability of baseline. The earlier recommendations regarding stability all apply here, with one addition. If first-phase data are unavailable or excessively variable, and if treatment must begin, a design might be used that does not require a baseline (e.g., an alternating treatments design).

A final consideration is the trend in baseline. When the following phase is expected to produce increases in the data, a falling or flat baseline is desirable. When deceleration is expected, rising or flat trends are beneficial. These are not rigid rules, however. A slowly rising baseline may be adequate if treatment is expected to increase it substantially. Again, these methodological suggestions coincide closely with good clinical judgment. If the client is already improving maximally, then the therapist should wait before beginning treatment.

Once again, these considerations actually apply to any phase in a within-series strategy. The logic of simple and complex phase changes is the same whether one is going from A to B or from C to D.

Implementing the second phase. To begin with, is there a variable that needs to be controlled first? For example, could any effects be due simply to, say, encouragement to change and not to the specific treatment? If this is highly plausible, and especially if treatment is costly, difficult, aversive, or restrictive, the alternative treatment (e.g., encouragement) might be implemented first. This parallels good clinical decision making and may fit in with legal requirements, such as the initial use of the least restrictive alternative. If the less restrictive treatment does not work, there is still the option of implementing the full treatment (see below).

Another consideration in implementing treatment or any new phase is that it should begin in full force if possible. Gradual implementation might minimize apparent differences between phases. This is a difficult issue (Thomas, 1978), but violating this rule only makes positive findings less likely. Once found, clear results are not jeopardized.

When the second phase is implemented, only three outcomes are possible: no improvement, deterioration, or improvement. If there is no improvement, the clinician has three reasonable paths open, both clinically and methodologically. One is to wait to see if there is a delayed effect. A second option, also a common clinical step, is to try another treatment strategy. It is typically assumed that a phase producing no change can be with caution considered part of the previous phase (e.g., A=B / C). As phases are added, the plausibiliy of equivalence is jeopardized (e.g., an A=B=C=D=E / F / E / F design seems weak). The solution is to be had in systematic replication across clients (e.g., several A/F/A/F designs could be added to

the one above). Finally, treatment can be altered by adding or subtracting components (e.g., A=B / B+C / B), also a common clinical step.

If treatment produces deterioration, the clear course is to withdraw treatment. If the behavior once again improves, the clinician will have documented an iatrogenic effect of treatment, often itself a significant contribution to the field.

The final possible effect of the second phase is improvement, which opens three possible paths. First, the clinician can continue treatment through to a successful conclusion and store the resulting A/B design. When a similar case presents itself, a multiple baseline across persons can be attempted. This is an extremely useful option and will be discussed at length later in the article . Second, if the client has other similar problems or problem situations, apply the same treatment to them (again a multiple baseline). A final course of action is to withdraw the treatment or implement a treatment placebo. If improvement then slows, a treatment effect is more likely.

The use of withdrawal is so popular that many confuse this design option with all of single case methodology, so a more extended discussion is warranted. There are potential problems with the withdrawal of an apparently effective treatment. It raises ethical issues, client fee issues, potential client morale problems, and possible neutralization of subsequent treatment effects. Few data exist on the actual likelihood of these problems, however, and there are many important counterarguments to be made (e.g., Hersen & Barlow, 1976).

The issue of withdrawal relates in special ways to the practicing clinical environment. First, if the treatment is of unknown benefit, a withdrawal can avoid the unnecessary use of ineffective treatment. Physicians recognize this issue in the common practice of drug holidays (i.e., withdrawals) to assess the continued need for treatment. Second, withdrawals often present themselves naturally in treatment in the form of vacations, holidays, sickness, temporary treatment dropouts, and the like. These can then be incorporated into ongoing clinical evaluations by examining measurements taken during or after these periods but before reintervention. Unlike withdrawals determined by the clinician, however, natural withdrawals are more likely to reflect variables of importance to these measures. For example, deterioration following treatment dropout may be due to factors producing that very decision rather than to the withdrawal of treatment per se. Therefore, clinicians should specify reasons for natural withdrawals and stress greater caution and need for replication when presenting cases with natural withdrawals. Third, withdrawals need not be long and drawn out. The slight delay in treatment that they impose should be weighed against their clinical value. Fourth, a good rationale that will minimize client morale problems can usually be given. The rationale can be either absolutely honest (e.g., "You've been rather successful so far with this approach, but I'm not sure we still need to be following this course, so let's take a little breather and see where things go"), or they can be somewhat deceptive (e.g., giving the client the expectation that treatment is normally stopped now and that this often leads to even greater improvement). Such placebo rationales must be handled with care, of course, just as a placebo drug might be in medical practice. Fifth, withdrawals are often produced when turning to other issues. For example, the clinician may wish to spend a few weeks in reassessment of the client. While clinically valuable, this might constitute an attention placebo for a specific problem under treatment. This is a type of withdrawal, just as data taken during an initial assessment phase (which involves much more than mere baseline measurement) is thought of as baseline. Finally, withdrawals often have clear clinical benefit to the client. If behavior worsens, the client may become convinced that treatment is necessary and successful. If not, the client may see that the problem is now under control.

After withdrawal. If the clinician returns to the first condition following improvement on the second, three possible outcomes once again occur: deterioration, no change, or continued improvement. If the behavior deteriorates, the clinical and methodological course is clear: Reimplement the effective

treatment (e.g., an A/B/A/B). If the behavior shows continued and further improvement, several options are available. One option is simply to wait. As in any situation in which the behavior is already improving, there may be little reason to further intervene. Sometimes the behavior will soon stop improving or deteriorate, perhaps due to a short-lived carry-over effect from the second phase. If, however, the behavior keeps improving significantly, the clinician can allow the case to continue to a successful conclusion and store these data, waiting for a similar case. This sequence can then be repeated but with a longer or shorter initial phase as part of a multiple baseline across subjects. If the effect is subsequently replicated and order effects eliminated, nonreversible improvement due to treatment will have been documented. If improvement continues in the withdrawal phase, the same sequence can be followed with another of the client's problem behaviors or the problem behavior in another situation, again producing multiple baselines. If the continued improvement is not maximal, treatment can be reimplemented anyway. A subsequent increase in the rate of improvement would establish greater confidence in the treatment.

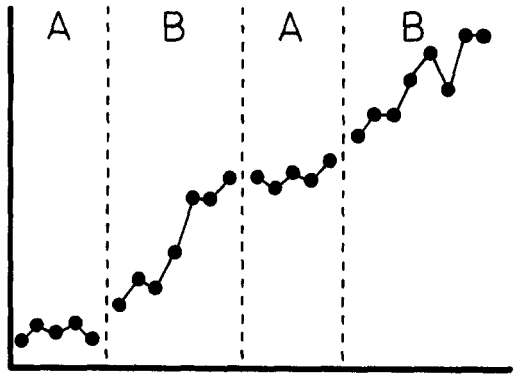
If no change is seen when the second phase is withdrawn (the behavior shows neither deterioration nor continued improvement), the options described above are open. The reimplementation option is particularly attractive. Some methodologists might be concerned over this advice, since the level of the behavior shown in baseline was not reattained in the return to baseline phase. This is a difficult argument for the clinician, since it implies that lack of maintenance of behavior change is a requirement in order to show treatment effects when using a withdrawal. Essentially, this would have clinicians document success by showing failure. Fortunately, it is the history of single case methodology, not its logic, that enables such a problematic argument to be made. For example, animal operant researchers (especially historically) have often seemed to assume that current behavior is primarily a function of immediately present environmental variables. Thus, behavior should be in one steady state when these variables are present and in another when they are not. This type of assumption pervades much of single case methodology, often to the detriment of its clinical uses.

The assumptions of the clinician are quite different. The clinician usually assumes that the current level of behavior is often a function of historical variables as much as current conditions. Greater improvement may be expected to be associated with treatment, but the actual level of behavior hopefully is maintained even when treatment is withdrawn (cf. Sidman, 1960, on transition states).

When these assumptions are applied to the logic of within-series strategies, it is apparent that deterioration (return to the previous level) is not required during withdrawal. If behavior improves faster during treatment than not, an effect is shown. It may be useful to regraph some data to underscore this. The top half of Figure 2, for example, shows an A/B/A/B sequence in which withdrawal produces less improvement but no clean reversals. The bottom half of the figure shows the same data calculated as difference scores from the trend in the previous phase (or same phase in the case of the first phase). When plotted in terms of improvement, a more classical pattern emerges.

There are many other ways in which the assumptions of the typical clinician overturn nonessentials of time series methodology as developed by operantly oriented psychologists and lead to new design options. For example, the notion of treatment phases as easily identifiable entities is jeopardized. A good deal of clinical work is done under 1 hour per week outpatient conditions. Sometimes, it is true, clinicians use this time to set up treatments that are obviously present throughout a specifiable time (e.g., a token economy for a noncompliant child). However, other clinicians (e.g., especially those working with adults) do not change obvious aspects of the client's environment outside of the clinical session itself. When, then, can treatment be said to be present? During that hour? That day? That week?

Ambiguity about the meaning of the word *phase* is not lethal to the clinical use of within-series strategies, but it does help open up new design options. It is not lethal, be-



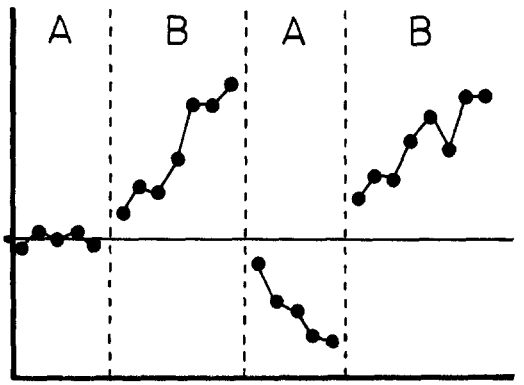


Figure 2. An example of the nondeterioration of behavior in the withdrawal phase of a within-series design, hypothetical data. (The lower graph is calculated in terms of improvement to highlight the control shown over the transitional state of behavior. The upper graph and lower graph both demonstrate experimental control despite nonreversibility; the lower is merely more obvious.)

cause (a) the effects of treatment often last well beyond the actual therapy hours, (b) any ambiguity about the nature of phases (e.g., Thomas, 1978) makes only robust effects visible, and (c) phases usually incorporate considerable lengths of time. Thus, ambiguity about what is in one phase or in another is not a major threat to the internal validity of any clear effects actually obtained.

The design element opened up by this issue is the periodic treatments element (Hayes & Nelson, Note 1). The notion is that a consistent relationship between the periodicity of treatment and the periodicity of behavior change can demonstrate therapeutic effects. This relationship can only be shown when the frequency of behavioral measurement far exceeds the frequency of treatment sessions. An example may show the principle. The top half of Figure 3 shows the hypothetical record of positive social interactions self-recorded daily by a client. Arrows on the abscissa show days when the client saw a psychotherapist for 1-hour insight-oriented therapy sessions. Since the treatment sessions occur at varying intervals, and periods of improvement only follow them, these changes are likely due to treatment. The

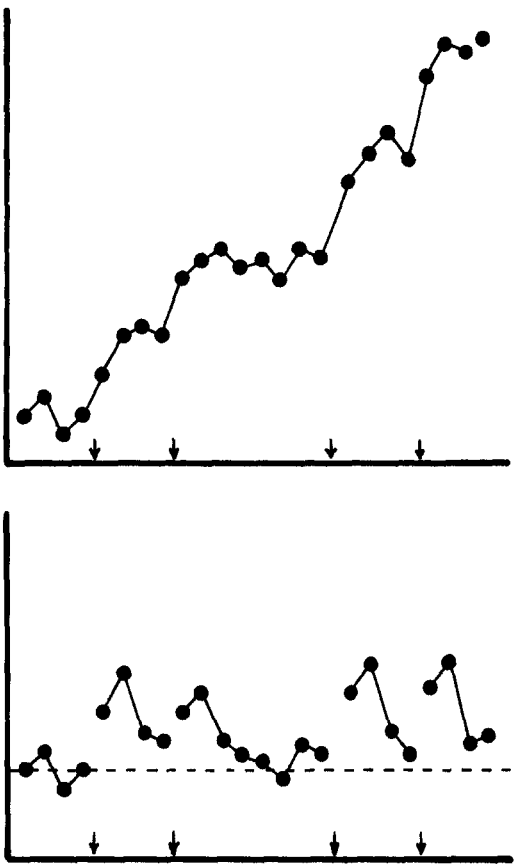


Figure 3. The periodic treatments effect is shown on hypothetical data. (Data are graphed in raw data form in the top graph. Arrows on the abscissa indicate treatment sessions. This apparent B-only graph does not reveal the periodicity of improvement and treatment as well as the bottom graph, where each two data points are plotted in terms of the difference from the mean of the two previous data points. Significant improvement occurs only after treatment. Both graphs show an experimental effect; the lower is merely more obvious.)

bottom half of the figure presents the data in difference score form, which draws this out even further. These data do not show what about the treatment produced the change (any more than an A/B/A design would). It may be therapist concern or the fact that the client attended a session of any kind. These possibilities would then need to be eliminated. For example, one could manipulate both the periodicity and nature of treatment. If the periodicity of behavior change was shown only when a particular type of treatment was in place, this would provide evidence for a more specific effect.

The periodic treatments element has apparently not been used in a published study (this is the first published description of the design element), although some of my own cases have shown clear examples of such periodicity (e.g., Hayes, Note 2). The major point is that clinical assumptions seem to lead to different design elements than those generated by the animal laboratory. It is possible that new developments in single case designs will occur as the needs and assumptions of practicing clinicians have more of an effect on the methodology itself.

Between-Series Strategies

In contrast with the within-series elements, in which changes within a series of data points are compared, the between-series strategies compare two or more series of data points across time. The comparisons are repeatedly made between these series. There are two basic types of pure between-series elements: the alternating treatments design and the simultaneous treatments design.

The Alternating Treatments Design Element

The logic of the alternating treatments design (Barlow & Hayes, 1979) is based simply on the rapid and random (or semi-random) alternation of two or more conditions, in which there is one potential alternation of condition per measurement opportunity. Since a single data point associated with one condition may be preceded and followed by measurements associated with other conditions, there is no opportunity to estimate stability, level, and trend within phases. Rather, these estimates are obtained within conditions, by collecting measurements associated with a condition each into a separate series. If there is a clear separation between such series, differences among conditions are inferred. For example, suppose a clinician wishes to examine the relationship of therapist self-disclosure to client self-disclosure. At the beginning of some sessions (randomly determined), the therapist self-discloses; in the other sessions, no self-disclosure is used. Tape recordings of the sessions are rated (see Figure 4), with results demonstrating that therapist self-disclosure increases client self-disclosure. Note that the comparison is made purely between series. The general upward trend in each condition is not analyzed and may be due to extraneous factors, but the major comparison is still sound. Thus, the alternating treatments strategy is viable even if within-series trends are extreme or are changing rapidly (e.g., in learning situations or with maturational phenomena).

One could think of this as an extremely rapid A/B/A (cf. Campbell & Stanley's, 1963, discussion of an "equivalent time samples" design), but they differ significantly. Not only are the estimates of variability and source of treatment comparisons different,

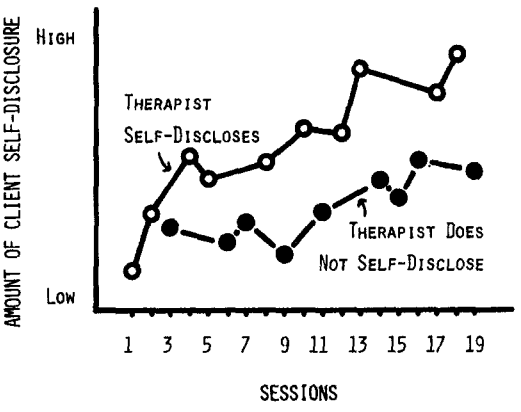


Figure 4. An example, using hypothetical data, of the alternating treatments design element. (The clear difference between the two conditions shows that more client self-disclosure is produced when the therapist self-discloses. The overall increase across time is not analyzed without the addition of other design elements; e.g., A phases before and after.)

but this design also minimizes order effects (by random sequencing) and can incorporate three or even more conditions into a single comparison sequence (see Barlow & Hayes, 1979).

This design strategy is often combined with other design elements (e.g., a baseline), though it is not required. It is particularly useful for the comparison of two or more treatments or when measurement is cumbersome or lengthy (e.g., an entire MMPI). Only four data points are absolutely needed (two in each condition). Each data point may incorporate many treatment sessions; the rapid alternation refers only to the rate of treatment alternation relative to the rate of measurement. On the other extreme, alternations might be made several times per session (e.g., Hayes, Hussian, Turner, Grubb, & Anderson, Note 3).

This design is also valuable when difficult assessment decisions are presented. Suppose, for example, that a client is presenting with social deficits. The clinician may have a difficult time determining if the client is more likely to respond to anxiety management procedures or social skills training procedures. Rather than guess, the clinician might do both in an alternating treatments fashion. The better treatment may quickly be revealed, and all treatment effort could then go in this direction.

The Simultaneous Treatment Design Element

The only other true between-series element is the simultaneous treatment design (Browning, 1967). It requires the simultaneous presence of two or more treatments. Since the treatments are truly available simultaneously, the client controls which treatment is actually applied (much as in a concurrent schedule design in animal operant work). Thus, a true instance of this design can only measure treatment preference, not treatment effectiveness. Apparently only one example (Browning, 1967) exists in the applied literature. (As for Kazdin & Hartmann, 1978, and McCollough, Cornell, McDaniel, & Meuller, 1974, see Barlow & Hayes, 1979.) However, many current applied questions (e.g., about the relative aversiveness or restrictiveness of treatments) are issues of preference, and the simultaneous treatments design might be of real use in these situations.

Combined-Series Strategies

Several design elements in time series experimentation borrow from both of the previously described strategies. These combined-series elements utilize coordinated sets of comparisons made both between and within series of measurements.

The Multiple Baseline

Undoubtedly the most familiar combined-series element is the multiple baseline. Its logic is intended to correct for major deficiencies of a simple phase change (say, an A/B). In an A/B, any changes between the two phases could be due to coincidental extraneous events: maturation, cyclical behavior, baseline assessment, and so on. The mul-

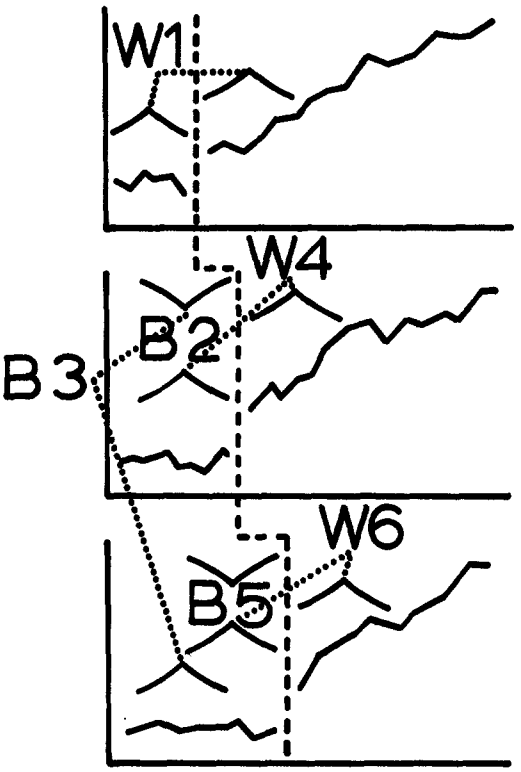


Figure 5. The types of comparisons made in a multiple baseline. (W = a within-series comparison, and B = a between-series comparison. The numbers show the usual sequence of comparisons).

tiple baseline solves these problems by replicating the A/B but with different lengths of baseline for each replication (a strategy that controls for the amount of baseline assessment or mere maturation) and with the actual time of the phase change arbitrarily altered (to reduce the possibility of correlated extraneous events).

As is shown in Figure 5, a typical multiple baseline allows several comparisons. Some are identical to those made in a simple phase change, whereas others are between-series comparisons, examining patterns within an unchanged series compared to phase changes in other series.

A multiple baseline can be done with a similar behavior in two or more clients (across people), two or more behaviors in one client (across behaviors), or with a behavior in two or more settings in one person (across settings). The specific phase changes, however, must be the same—the same first condition must yield to the same second condition—since it is alternative explanations for a specific phase change effect that are being controlled.

The label *multiple baseline* is something of a misnomer. The logic of the comparison applies to any set of phase changes so arranged, whether or not there is a baseline present. For example, a series of B/C phase changes could easily be arranged into a multiple baseline (*multiple phase change* would actually be a clearer term). Sometimes it is used sequentially; for example, the sequence A/B/C (with A/C/B to control for order effects) can be put into a type of multiple baseline, as shown in the top half of Figure 6. This arrangement is problematic, since the third phase is introduced after equal second-phase lengths in each series (not controlling for sudden maturational or for phase length effects in the B/C comparison). A better sequential multiple baseline is shown in the bottom half of Figure 6.

No absolute rule can be given about the number of phase shift replications required between series in a multiple baseline element. The logic of the maneuver applies as well to a single replication as to several; it is simply that each additional series strengthens our confidence that much more. Thus, the clinician should not feel as though the element is useless when only two series are compared, though more are desirable. The same can be said about the differences in initial phase length. If one series has an

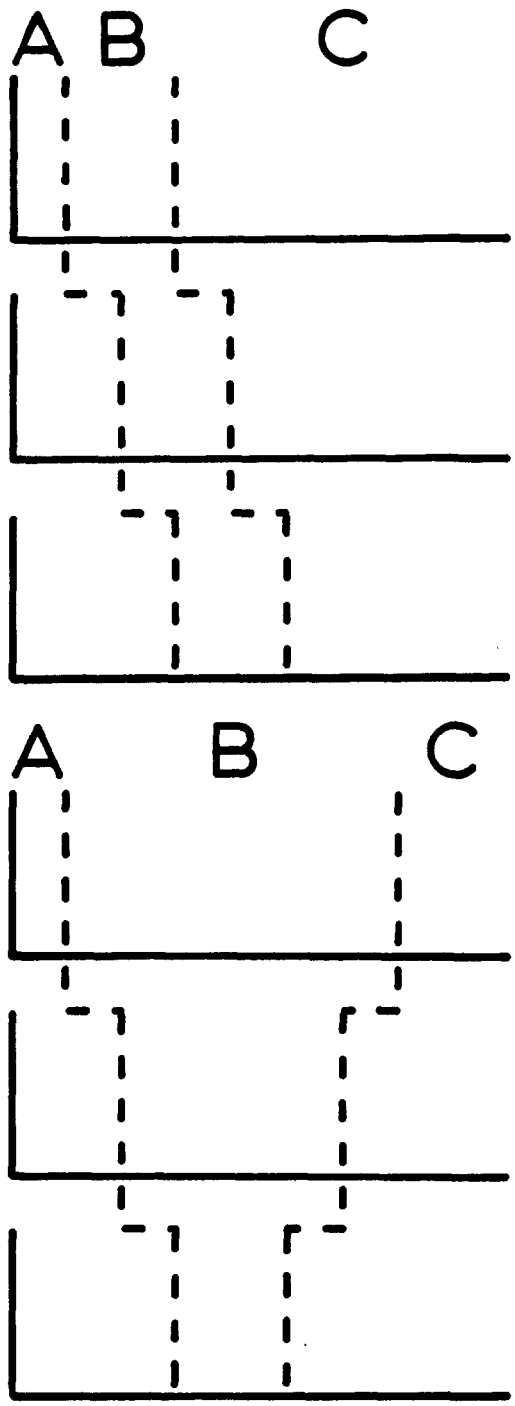


Figure 6. An example of a weak (top) and strong (bottom) arrangement in a sequential multiple baseline.

initial phase only slightly longer or shorter than the other, this is less satisfactory than if there are large differences.

Much has been made of the need to avoid the multiple baseline when the specific series are interdependent (e.g., Kazdin & Kopel, 1975). If a phase shift in a multiple baseline is accompanied by behavior change not only within series but also between series, it is difficult to distinguish uncontrolled effects from true treatment effects. For example, in a multiple baseline across behaviors, changes in one behavior may produce changes in another, because of actual processes of response generalization caused by treatment. Typically, this is not a problem in the use of multiple baseline elements so much as it is an opportunity to study generalization effects. Thus, the clinician in this situation could immediately embark on a new design (e.g., withdraw treatment and see if both behaviors stop improving), which would document that the multiple effects are actually being caused by treatment, an important contribution. Further, if several series are being compared, some interdependence can be tolerated (e.g., Hayes & Barlow, 1977; Hersen & Bellack, 1976) without undoing the design (Kazdin, 1980).

The opportunity to use the multiple baseline element in clinical practice is very large. Multiple baselines often form naturally across behaviors due to the tendency for practicing clinicians to tackle subsets of problems sequentially rather than all at once. Multiple baselines across settings are less common but also naturally occur when clinicians treat problem behavior shown in one specific condition first rather than treating the problem all at once (e.g., Hayes & Barlow, 1977).

The multiple baseline across people is probably one of the clearest examples of natural design elements that arise in clinical practice. Nothing could be more natural to clinical work than an A/B. To form a multiple baseline, all the clinician need do is save several of these with similar problems and the same treatment. Individual clients will inevitably have differing lengths of baseline, often widely so, due to case complexities or to matters of convenience. Thus, sequential cases usually lead to multiple baseline across people.

Some of the earliest applied literature on the multiple baseline (e.g., Baer, Wolf, & Risley, 1968) stated that multiple baselines across persons should always be done at the same time in the same setting with the same behavior. Saving cases, with perhaps periods of months or even years separating each, violates this rule, but fortunately the logic of the strategy does not really require it. If the time of the phase shift differs in real time from client to client, it is unlikely that important external events could repeatedly coincide with the phase changes. The control exerted by the different lengths of baseline remain.[2]

There is a potential difficulty, which was touched on in the earlier discussion of natural withdrawals. If the clinician is allowing the case itself to determine the exact length of baseline, there is the danger that the same factor which indicated that it is time to change phases is correlated with processes that produce behavior change. The main practical protections against this difficulty are replication (including several cases in natural multiple baselines) and information (reporting why the phase was changed for each client). If reasons for changing phases vary from client to client, it is unlikely that a third variable consistently produced changes in the second phase.

It is also essential that clinicians report all cases attempted, not just those showing the desired effect. If the effect is not seen in some of the cases, the clinician should attempt to find out why; indeed this seems required by good clinical practice. A careful examination of possible differences between individuals accounting for variable results may lead to treatment solutions for nonresponsive clients. Data showing subsequent response would increase our knowledge about mechanisms producing change and about boundary conditions of a given treatment.

---

[2] A minor weakness is the fact that the events in real time that might have produced the phase change are not present in the other series. If effects are clear, this need not be a concern, since a series of such coincidences is still unlikely.

The multiple baseline across cases also provides a home for those cases in which just treatment is given (B only) and in which treatment is never given (baseline-only control; see below). As anchors in a series of cases arranged into a multiple baseline across subjects, such cases can provide evidence of the effectiveness of treatment even when no baseline is taken (B only), thus controlling for an unlikely order effect due to A or of the likelihood of change when no treatment is given (baseline-only control).

Crossovers

This maneuver (drawn from similar group-comparison approaches; see Kazdin, 1980) is based on two concurrent phase changes, one in reverse order of the other. For example, one subject may experience a B/C sequence; the other, a C/B. By changing phases at the same time, this strategy equalizes alternative sources of control that might have produced an apparent phase change effect (e.g., maturation, phase length). Since these sources are equalized, consistent within-series effects in the two series (e.g., if B > C in both cases) provide evidence for the comparison. The controls are not strong, however (e.g., order effects are weakly dealt with), so the entire crossover should be replicated at least once with other clients. Some of these same issues apply to the true reversal, which also is a combined series element (see Leitenberg, 1973).

The Baseline-Only Control

Many times problems are repeatedly assessed but never treated. This may be done deliberately (e.g., assessing those on a waiting list) or serendipitously (e.g., assessing a problem behavior that is never treated, because the client moves away). Whatever else has been done, these data can be examined as a type of combined- or between-series comparison (e.g., Brownell, Hayes, & Barlow, 1977; Hayes & Cone, 1977). The logic of this comparison is identical to the between-series comparisons made in a multiple baseline design (see Figure 5). Changes occurring elsewhere and not in the baseline-only control series are more likely to have been produced by treatment (cf. Campbell & Stanley's, 1963, equivalent time samples design).

Issues in the Use of Time Series Design Tools

The purpose of the present article is to provide an overall framework for present single case design tools and to point out how they might fit into evaluations of actual clinical practice. If these tools are to be used by large numbers of practicing clinicians, many specific problems need to be solved (e.g., development of practical measurement tools, methods of specifying of treatment activities), but the most important problem is one of overall approach. By repeatedly emphasizing design elements rather than complete designs, the present organization is meant to encourage creative evaluations in actual clinical decision making. These are not static tools. It is quite possible to devise designs without names, designs in which many of the elements mentioned in the article are combined. As the clinician approaches each case, questions arise that require answers on clinical grounds. If the clinician is aware of available design options, some time series strategy is almost always available that fits closely with the logic of clinical decision making itself.

Table 1 presents some clinically important questions and examples of the various design elements useful in that situation. Within any row of this table, various elements can be combined to address a given clinical question. As different questions arise, different elements can be used (draw from different rows).

Another major stumbling block in the use of time series design tools in clinical practice is the historical status of the division between practice and research. At first glance, the distinction between research and treatment is clear cut and easily applied. Clinicians who have not used the type of approach advocated here often very easily define research and treatment in terms of their apparent structure (Hayes, Note 4), such as (a) Did the clinician collect systematic data? (b)

Table 1
Examples of the Use of Design Elements to Answer Specific Types of Clinical Questions

Clinical question	Design type		
	Within series	Between series	Combined series
Does a treatment work?	A/B/A/B/. . . B/A/B/A/. . . A/B (see combined designs) Periodic treatments design Changing criterion design	Alternating treatments (comparing A and B)	Multiple baseline across settings, behaviors, or persons comparing A and B Replicated crossovers (comparing A and B)
Does one treatment work better than another, given that we already know they work?	B/C/B/C/. . . C/B/C/B/. . .	Alternating treatments (comparing B and C)	Replicated crossovers (comparing B and C) Multiple baselines (comparing B and C and controlling for order)
Does one treatment work, does another work, and which works better?	A/B/A/C/A combined with A/C/A/B/A	Alternating treatments (comparing A and B and C)	Multiple baseline (comparing A and B and C and controlling for order)
	Or combine any element from Row A with any element from Row B		
Are there elements within a successful treatment that make it work?	B / B+C / B B+C / B / B+C C / B+C / C B+C / C / B+C	Alternating treatments (comparing, for example, B and B+C)	Multiple baseline (comparing B and B+C, and C and B+C) Replicated crossovers (comparing B and B+C, and C and B+C)
Does the client prefer one treatment over another?		Simultaneous treatments (comparing B and C)	
Does a treatment work, and if it does, what part of it makes it work?	Combine any elements from Rows A or C with any element from Row D		
What level of treatment is optimal?	Ascending/descending design B/B′/B/B′	Alternating treatments (comparing B and B′)	Multiple baseline (comparing B and B′ and controlling for order) Replicated crossovers (comparing B and B′)

Were the variables producing the impact systematically analyzed? (c) Were the results of this endeavor presented or published? The presence of any one of these is likely to lead to the endeavor's being termed "research." The consequences of this can be dramatic. We have generated a large number of protections in research with human subjects. It is possible, however, to use the structure of research to perform the function of treatment. This is treatment evaluation or empirical clinical practice (Jayaratne & Levy, 1979).

The ethical questions posed by treatment as opposed to treatment evaluation seem very similar. Indeed, the effects of evaluations of the sort described here seem beneficial on two grounds (Levy & Olson, 1979). First, the attempt to evaluate treatment is likely to contribute to clinical effectiveness by increasing feedback to the clinician and client alike, by increasing the clinician's involvement in the case, and by increasing information available about the client's response to treatment. Second, by increasing the knowledge base in the field more generally, such an approach would make successful treatment of others more likely.

Nevertheless, practicing clinicians (and society more generally) often make a distinction between treatment and evaluation based on mere appearance. In particular, evaluation is often grouped with research rather than with treatment per se. The effect of this is to discourage empirical clinical practice, since it leads to a number of additional protections beyond that required in the treatment environment itself. Unless this process is resisted (e.g., by not submitting routine clinical evaluation to human-subjects committees), strong negative pressure is put on the practicing clinician to avoid systematic evaluation.

Additional problems remain (Hayes, 1980; Levy & Olson, 1979; Thomas, 1978). For example, if the approach advocated here were adopted, a flood of information could emerge from the many thousands of practicing clinicians. Where would it be put? Who would publish it? Would it be simple-minded research anyway? Multiple case manuscripts might be a partial solution; a clearing-house-type arrangement might also be of aid, but it clearly would strain current information-handling systems.

Another problem is the importance of compromises forced by the clinical environment. There are a number of them (Thomas, 1978), although most seem soluble. The major solution is the same as that for most difficulties in time series designs more generally: replication. Only with the enormous resources provided by practicing clinicians does this advice seem practical. Without them, the external validity of single case work, which emerges only from replication (Hersen & Barlow, 1976), has little chance of full demonstration or analysis.

This, then, is the situation. Practicing clinicians are essential to the development of our knowledge base in clinical psychology, and time series experimentation seems fully applicable to the clinical environment. Indeed, the resources needed to repeatedly replicate single case experimentation are available only by including practicing clinicians. If combined, these needs, abilities, and resources could create a true revolution in clinical psychology. The question is, will they be?

Reference Notes

1. Hayes, S. C., & Nelson, R. O. *Realistic research strategies for practicing clinicians.* Workshop presented at the meeting of the Association for Advancement of Behavior Therapy, San Francisco, December 1979.
2. Hayes, S. C. *A component analysis of flooding relief.* Paper presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.
3. Hayes, S. C., Hussian, R., Turner, A. E., Grubb, T., & Anderson, N. *The nature, effect, and generalization of coping statements in the treatment of clinically significant anxiety.* Paper presented at the meeting of the Association for Advancement of Behavior Therapy, San Francisco, December 1979.
4. Hayes, S. C. *Ethical dilemmas of the empirical clinician.* Paper presented at the meeting of the Association for Behavior Analysis, Dearborn, Michigan, May 1979.

References

Baer, D. M., Wolf, M. M., & Risley, T. R. Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1968, *1*, 91–97.

Barlow, D. H. Behavior therapy: The next decade. *Behavior Therapy*, 1980, *11*, 315–328.

Barlow, D. H., & Hayes, S. C. Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis*, 1979, *12*, 199–210.

Barlow, D. H., & Hersen, M. Single-case experimental designs: Uses in applied clinical research. *Archives of General Psychiatry*, 1973, *29*, 319–325.

Becker, I. M., & Rosenfeld, J. G. Rational–emotive therapy—A study of initial therapy sessions of Albert Ellis. *Journal of Clinical Psychology*, 1976, *76*, 12–26.

Bernard, C. *An introduction to the study of experimental medicine*. New York: Dover, 1865.

Box, G. E. P., & Jenkins, G. M. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1976.

Brownell, K. E., Hayes, S. C., & Barlow, D. H. Patterns of appropriate and deviant arousal: The behavioral treatment of multiple sexual deviations. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 1144–1155.

Browning, R. M. A same-subject design for simultaneous comparison of three reinforcement contingencies. *Behaviour Research and Therapy*, 1967, *5*, 237–243.

Browning, R. M., & Stover, D. O. *Behavior modification in child treatment: An experimental and clinical approach*. Chicago: Aldine-Atherton, 1971.

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

Chassan, J. B. *Research design in clinical psychology and psychiatry*. New York: Appleton-Century-Crofts, 1967.

Chassan, J. B. *Research design in clinical psychology and psychiatry (2nd ed.)*. New York: Irvington, 1979.

Garfield, S. L., & Kurtz, R. Clinical psychologists in the 1970s. *American Psychologist*, 1976, *31*, 1–9.

Glass, G. V., Wilson, V. L., & Gottman, J. M. *Design and analysis of time-series experiments*. Boulder: University of Colorado Press, 1975.

Gottmann, J. M., McFall, R. M., & Barnett, J. T. Design and analysis of research using time series. *Psychological Bulletin*, 1969, *72*, 299–306.

Hartmann, D. P., & Hall, R. V. The changing criterion design. *Journal of Applied Behavior Analysis*, 1976, *9*, 527–532.

Hayes, S. C. A review of Jayaratne & Levy's "Empirical clinical practice." *Behavioral Assessment*, 1980, *2*, 306–308.

Hayes, S. C., & Barlow, D. H. Flooding relief in a case of public transportation phobia. *Behavior Therapy*, 1977, *8*, 742–746.

Hayes, S. C., & Cone, J. D. Reducing residential electrical energy use: Payments, information, and feedback. *Journal of Applied Behavior Analysis*, 1977, *10*, 425–535.

Hersen, M., & Barlow, D. H. *Single case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press, 1976.

Hersen, M., & Bellack, A. S. A multiple baseline analysis of social skills training in chronic schizophrenics. *Journal of Applied Behavior Analysis*, 1976, *9*, 527–532.

Holt, R. R. (Ed.). *New horizons for psychotherapy: Autonomy as a profession*. New York: International Universities Press, 1971.

Jayaratne, S., & Levy, R. L. *Empirical clinical practice*. New York: Columbia University Press, 1979.

Kazdin, A. E. Methodological and interpretive problems of single-case experimental designs. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 629–642.

Kazdin, A. E. *Research design in clinical psychology*. New York: Harper & Row, 1980.

Kazdin, A. E., & Hartmann, D. P. The simultaneous-treatment design. *Behavior Therapy*, 1978, *9*, 912–922.

Kazdin, A. E., & Kopel, S. A. On resolving ambiguities of the multiple-baseline design: Problems and recommendations. *Behavior Therapy*, 1975, *6*, 601–608.

Kelly, E. L., Goldberg, L. R., Fiske, D. W., & Kilkowski, J. M. Twenty-five years later. *American Psychologist*, 1978, *33*, 746–755.

Kratochwill, T. F. *Single subject research: Strategies for evaluating change*. New York: Academic Press, 1978.

Leitenberg, H. The use of single case methodology in psychotherapy research. *Journal of Abnormal Psychology*, 1973, *82*, 87–101.

Leitenberg, H. Training clinical researchers in psychology. *Professional Psychology*, 1974, *5*, 59–69.

Levy, L. H. The skew in clinical psychology. *American Psychologist*, 1962, *17*, 244–249.

Levy, R. L., & Olson, D. G. The single subject methodology in clinical practice: An overview. *Journal of Social Service Research*, 1979, *3*, 25–49.

Mahoney, M. J. Experimental methods and outcome evaluation. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 660–672.

McCollough, J. P., Cornell, J. E., McDaniel, M. H., & Meuller, R. K. Utilization of the simultaneous treatment design to improve student behavior in a first-grade classroom. *Journal of Consulting and Clinical Psychology*, 1974, *42*, 288–292.

Meehl, P. E. A scientific, scholarly, non-research doctorate for clinical practitioners: Arguments pro and con. In R. R. Holt (Ed.), *New horizons fpr psychotherapy: Autonomy as a profession*. New York: International Universities Press, 1971.

Meltzoff, J., & Kornreich, M. *Research in psychotherapy*. New York: Atherton, 1970.

Millon, T., & Diesenhaus, H. I. *Research methods in psychopathology*. New York: Wiley, 1972.

Nelson, R. O. Realistic dependent measures for clinical use. *Journal of Consulting and Clinical Psychology*, 1981, *49*, 168–182.

Paul, G. L. Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969.

Peterson, D. R. Need for the doctor of psychology degree in professional psychology. *American Psychologist*, 1976, *31*, 792–798.

Raush, H. L. Naturalistic method and the clinical approach. In E. P. Willems & H. L. Raush (Eds.),

Naturalistic viewpoints in psychological research. New York: Holt, Rinehart & Winston, 1969.

Raush, H. L. Research, practice, and accountability. *American Psychologist*, 1974, *29*, 678–681.

Rogers, C. R. Some new challenges. *American Psychologist*, 1973, *28*, 379–387.

Shakow, D. What is clinical psychology? *American Psychologist*, 1976, *31*, 553–560.

Sidman, M. *Tactics of scientific research.* New York: Basic Books, 1960.

Svenson, S. E., & Chassan, J. B. A note on ethics and patient consent in single-case design. *Journal of Nervous and Mental Disease*, 1967, *145*(3), 206–207.

Thelen, M. H., & Ewing, D. R. Roles, functions, and training in clinical psychology: A survey of academic clinicians. *American Psychologist*, 1970, *25*, 550–554.

Thomas, E. J. Research and service in single-case experimentation: Conflicts and choices. *Social Work Research and Abstracts*, 1978, *14*, 20–31.

Zaro, J. S., Barach, R., Nedelmann, D. J., & Dreiblatt, I. S. *A guide for beginning psychotherapists.* Cambridge, England: Cambridge University Press, 1977.

Received September 2, 1980 ■