Multiple Baseline Designs

Leonard L. LaPointe
Veterans Administration Hospital
Gainesville, Fl

Evaluations of therapeutic techniques in aphasia which meet the rigorous criteria of the scientific community have been as elusive as the proverbial emission-free exhaust system.

Today, we have been invited to serve, not as converters, but as catalysts in our attempt to review some traditional approaches to research design; as well as to highlight the basic concepts of some other approaches.

As has been pointed out, the traditional experimental group, control group, statistical analysis method of research, has some disadvantages when applied to efficacy of treatment questions. Psychiatry and clinical psychology are as aware of these problems as we are, and perhaps have grappled with them longer. Matching large groups of patients with similar symptomatology is often very difficult, even if one can afford the costs of gathering data, following subjects, and analyzing the data. The ethical considerations of witholding treatment have proved problematic and, rightly or wrongly, have inhibited and in some cases prevented research.

An alternative is single case experimentation, which I might emphasize, should be viewed as an addition and a supplement to traditional strategies in pursuing research questions; but is particularly suited to intervention issues.

The theoretical and logical aspects of single case experimental design have been outlined well by Herson and Barlow (1976) and their influence permeates this discussion.

Dr. Davis has discussed quite thoroughly the use of sequential reversal or withdrawal designs. There are times, however, when A-B-A designs are not appropriate or are not feasible. Another method for demonstrating the controlling effects of therapeutic variables on subjects is known as the "multiple baseline" technique.

The purpose of this paper is to describe the multiple baseline design, along with some of its nuances, and extend the discussion to the issue of generalization to non-trained items in aphasia. Table 1 is an outline of some uses of multiple baseline designs.

Table 1. Multiple Baseline Uses

1. When withdrawal or reversal is inappropriate
   - A. Practical Limitations
     (Time) (Carry over effects)
   - B. Ethical considerations
   - C. No staff cooperation
2. To study generalization

First, there are occasions when withdrawal or reversal are not appropriate.

Several practical limitations may prevent an A-B-A design from being carried out. The most obvious is not enough time to institute two or more withdrawal phases. In-patients are discharged and are not always available for continued treatment on an outpatient basis. Also, carryover or momentum effects might continue across adjacent phases of the study. This is particularly true in drug studies in which active agents may persist physiologically over long periods of time.

Certain ethical considerations also would preclude the withholding of a treatment variable. This argument is not so compelling in aphasia since we are being hounded to provide data, as it would be in, for example, dealing with self- or other-destructive behaviors. If a treatment variable were effective in reducing murder, for example, withholding treatment to prove that the rate of this destructive behavior increases when not treated, would, at the very least, be in poor taste.

Even if a particular behavior does not have destructive effects, and the beneficial effects are assumed if not proven, researchers may not obtain sufficient staff cooperation to carry out withdrawal of treatment. Teachers, parents, hospital staff, and patients themselves are sometimes less skeptical than the scientific community; and are reluctant to jeopardize therapeutic gains, especially when a disorder is severe, has a lengthy history, and previous attempts at remediation have failed.

Another major use of multiple baseline designs is that they can be used to study the issue of carryover, generalization, or improvement on non-trained items, that bogey-man (or rather bogey-person) that lurks outside every clinic door.

Multiple Baseline Designs: Variations on a Theme

Three types of experimental strategies have been used by applied clinical researchers when withdrawals or reversals are not feasible; and they include multiple baseline, multiple schedule, and concurrent schedule designs.

These designs have not been used very extensively in any of the clinical literature, and in aphasia, their use is rare.

The operant literature, however, reveals a growing indication of the popularity of this strategy. Just a cursory glance at a recent Cumulative Index of JABA (the Journal of Applied Behavioral Analysis) reveals multiple baseline designs used in classroom studies; in a study of imitation; in delusional speech; in following instructions; in language training; in group verbal conditioning; in sentence training; in training problem solving; in training use of descriptive adjectives; in social skills training; in studying shoplifting; in modifying attention; and in teaching coin values.

The rationale for multiple baseline studies first appeared in the literature in 1968 by Baer, Wolf and Risley (1968) three researchers who have made important contributions to time-series methodology and who previously had enjoyed moderate success in Uncle Remus movies.

Behaviors

These authors point out that in the multiple baseline technique, a number of responses are identified and measured over time to provide baselines against which changes can be evaluated. With these baselines established, the researcher than applies an experimental or therapeutic variable to one of the behaviors, produces a change in it, and perhaps notes little or no change in the other baselines. Then, the researcher applies the same

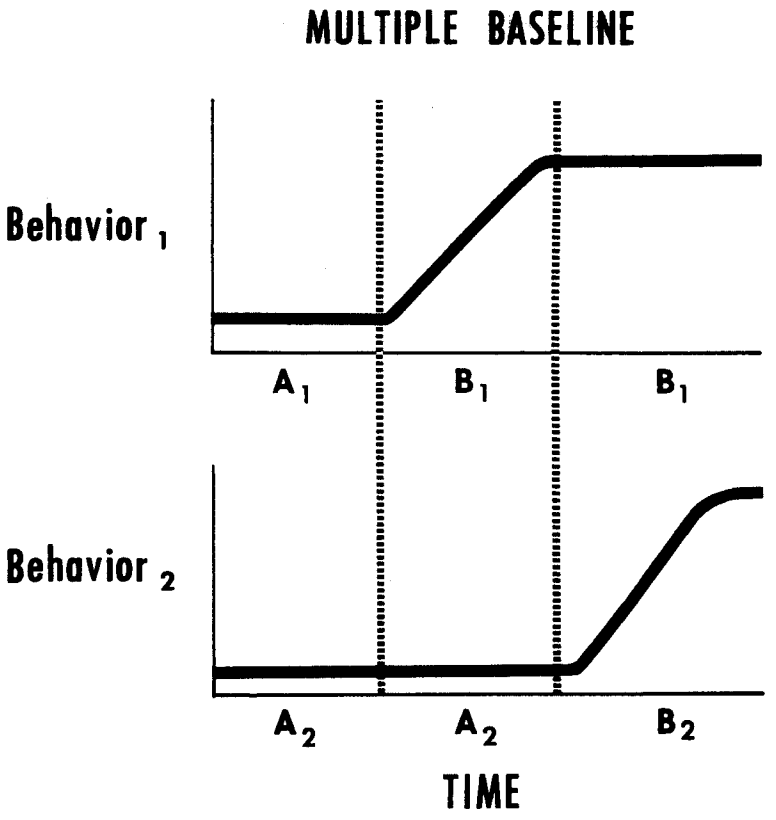

Figure 1. Miltiple baseline design: Two behaviors.

experimental or therapeutic variable to the second behavior and notes rate changes in that behavior.

This sequence can be visually appreciated in Figure 1. Notice that two behaviors are selected; baselines are measured; treatment is initiated on Behavior 1 (the B_1 phase) with a steep increase in performance. During the same time frame, baselines are continued on Behavior 2 (the A_2 phase). During the third time frame, treatment is initiated on Behavior 2 (the B_2 phase) with subsequent steep increase in performance on this variable.

In the first quadrant baselines are measured for o's, x's and squares, and stability is achieved. If you want to apply the example to aphasia think of the o's, x's and squares as three communication tasks; for example, matching, verbally repeating and writing ten verbs.

In the second quadrant, treatment is initiated on the o's only (matching); while basleines are continued on x's and squares. In the third quadrant treatment is begun on the x's (verbal repetition) with no treatment on the squares. And finally, treatment is initiated on the squares (writing).

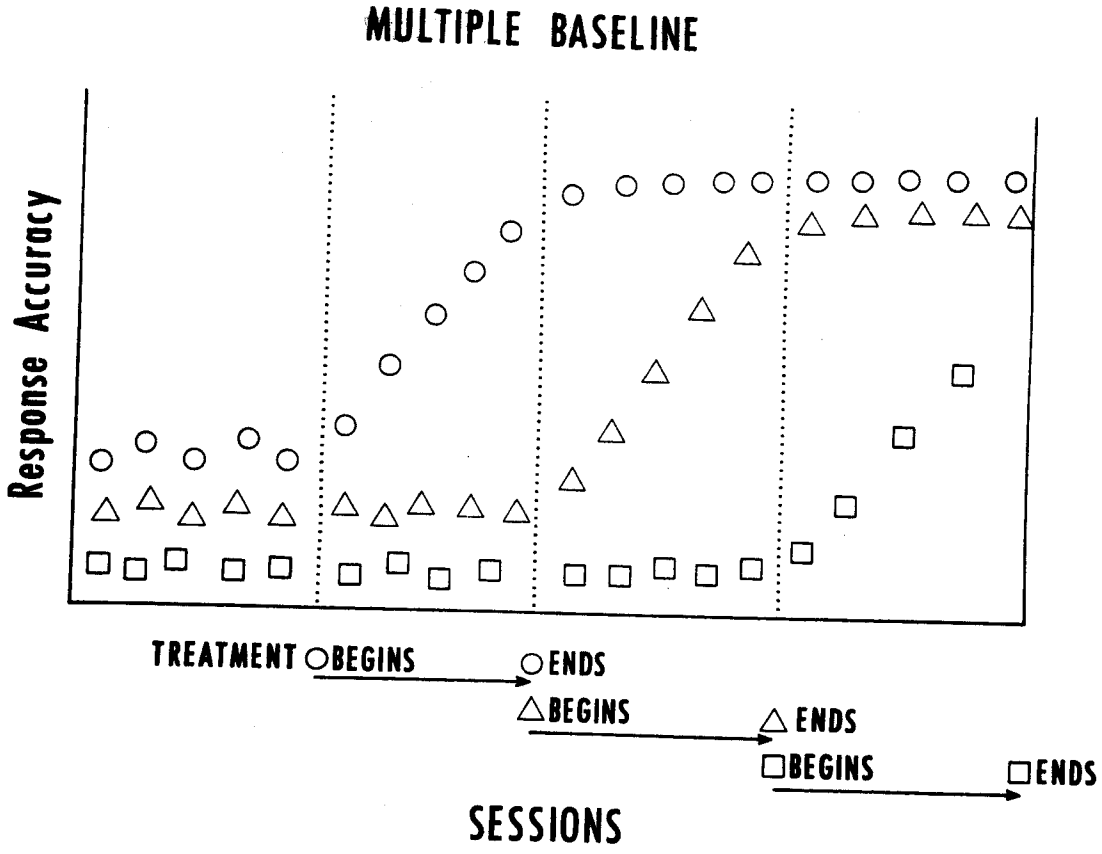

Figure 2. Multiple baseline design: Three behaviors with time-lagged treatment initiation.

Baseline and subsequent treatment interventions for each targeted behavior can be viewed as separate A-B designs, with each A phase further extended for each of the succeeding behaviors until the treatment variable is finally applied.

The researcher is assured that his treatment variable is effective when a change in rate appears after its application, while the rate of untreated behaviors remains relatively constant.

These are some of the basic principles underlying the multiple baseline design across behaviors. This design is applicable across subjects and across settings as well.

Subjects. In a multiple baseline design across subjects, a selected treatment is applied in sequence across matched subjects, presumably exposed to identical environmental conditions.

This strategy (across subjects) can be found in some of the classroom literature, and Hersen and Barlow (1976) cite the example of the sequential effects of a particular style of after-school tutoring on three students who were receiving D's and F's in a French class.

Settings. Finally, in the multiple baseline design across settings, a treatment variable is applied sequentially to the same behavior in the same subject across different and independent settings. Hersen and Barlow

(1976) cite an interesting example conducted in a summer camp which illustrates how both treatment and research considerations can be combined under naturalistic conditions (Allen, 1973). The subject in the particular example was a boy who presented a high frequency of bizarre verbalizations, primarily concerned with "penguins." During camping activities the child's high rate of bizarre verbalizations interfered with his developing good interpersonal relations both with peers and adults. An attempt was made to systematically reduce the high frequency of bizarre verbalizations in four separate camp activities (during an evening walk on the trail; in the dining hall; in the cabin and during a class).

During the first six days of baseline assessment in each setting no attempts were made to decrease the rate of bizarre verbalizations. Beginning on Day 7, a treatment procedure consisting of ignoring bizarre responses and attention to positive, non-bizarre, non-penguin verbalizations was implemented during one setting. Treatment was then applied in sequence under "time-lagged" conditions to the remaining three settings and resulted in a near zero rate of bizarre responses in all four settings. This is an excellent, though somewhat unusual, example of a successful multiple baseline design across settings.

Multiple Schedule Design

In this section I would like to review some of the variations of multiple baseline designs to suggest the versatility that exists in their application.

In the multiple baseline design across behaviors, separate and independent target behaviors are treated individually in sequence. In the multiple schedule design, though, the same behavior is treated differentially under varying stimulus conditions. For example, in the therapeutic context the different stimulus conditions for the same behavior might involve time differences, separate physical locations (which is very similar to "across settings"), different family members or different therapists.

An example closer to our own experience might be an attempt by two different clinicians to decrease a patient's perseverative responses, each clinician using a different treatment technique with the treatments instituted in sequence.

Concurrent Schedule Design

The final variation I shall discuss in the concurrent schedule design, and perhaps Table 2 will help clarify the differences. With this strategy, the subject is simultaneously exposed to different stimulus conditions, usually in a counterbalanced order. For example, one kind of concurrent schedule design might include attempts to decrease perseverative responses at home by family and attempts to decrease these behaviors during treatment by a clinician.

An advantage of this method is that the proportional efficacy of varying schedules (treatments) can be determined. Additionally, this design requires less time than do withdrawal, reversal, or other designs in which sequential treatment is used.

Table 2. Variations of Multiple Baseline Designs

MULTIPLE BASELINE (Behaviors):
Independent behaviors treated individually in sequence
MULTIPLE SCHEDULE:
Same behavior treated differentially under varying stimulus conditions in sequence
CONCURRENT SCHEDULE:
Same behavior treated simultaneously under different stimulus conditions

An example in Hersen and Barlow (1976) illustrates the use of this strategy to study reduction in "grandiose bragging" by a ten year old boy. Three reinforcement conditions ("praise, admonish, and ignore") were administered for three weeks by three teams of clinicians. In accordance with counterbalancing principles, each of the three teams administered all treatments, but in different orders for each of the three weeks.

The results showed dramatic decrease of "grandiose bragging" from the baseline phase and that the treatment condition of "ignore" was more successful than praising or admonishing. Interestingly, the authors note that, "Although it was not an objective of the study, the design and statistical analysis would enable an investigator to identify which staff might serve as the most effective behavior therapists to administer the selected treatment technique to the patient." (Hersen and Barlow, 1976, p. 108)

Generalization

An issue which is intimately related to multiple baseline designs and one which contributes to the restlessness of speech pathologists is that of generalization, or extension of stimulus control to non-trained items. The state-of-the-art of our understanding of generalization in aphasiology is at best shakey and at worst nauseating. We have established few principles about carryover to non-trained items and remarkably few studies have addressed the issue. In addition, the conclusions we can make from those which have addressed the issue are equivocal.

Carryover of improved comprehension skills following training has been both supported and challenged. Holland and Sonderman (1974) studied the effects of a program based on the Token Test for teaching comprehension skills to aphasic subjects and concluded that improved behavior in the trained task did not generalize to auditory comprehension skills in another setting.

In another report by Kushner and Winitz (1977), training and steady improvement of an aphasic patient's comprehension ability generalized across other modes to untrained verbal production skills.

Rosenbek, Green, Flynn, Wertz, and Collins (1977) reported performance of an anomic patient on sets of treated and untreated nouns. Generalization was reported on the untrained set by an increase in immediate correct responses and a reduction in number of necessary cues.

Other aspects of generalization have been studied or commented upon by Sidman (1971), Sidman et al. (1971), West (1973), Mohr et al. (1973) and LaPointe (1977). About the only postulate which emerges clearly from the research is that trained and untrained skills may be similar or dissimilar, depending on variations in stimulus mode, level of difficulty, and lexical, syntactic and phonologic factors. A few studies of generalization have been undertaken with children, generally involving articulation training. A study by Elbert, Shelton and Arndt (1967) is representative of these efforts.

We can learn from the operant literature, however, if we want to improve our appreciation of problems and advances in generalization. Stokes and Baer (1977) recently reviewed some 270 applied behavioral analysis studies relevant to generalization, and suggest that a technology of generalization programming is almost a reality. The need actively to program generalization, rather than passively to expect it as an outcome of certain training procedures, apparently is beginning to be realized.

Stokes and Baer (1977) have categorized techniques designed to assess or program generalization according to several general headings, as can be seen in Table 3.

Table 3. Techniques to Assess or Program Generalization

1. Train and hope
2. Sequential modification
3. Natural maintaining contingencies
4. Train sufficient exemplars
5. Train loosely
6. Loose trainly
7. Use indiscriminable contingencies
8. Program common stimuli
9. Mediate generalization
10. Train "to generalize"

(Stokes and Baer, 1977)

I will make no attempt to explain all of these strategies; some are self-explanatory; and I would refer you to the original article for details. But perhaps a few words about several of the concepts would be useful.

(1) Train and hope. This is the most frequent "method" reported. After a behavior change is effected through manipulation, any existent change across responses, settings, etc. may be documented or noted, but not necessarily pursued.

(2) Sequential modification. If generalization is absent or deficient, procedures are initiated to accomplish change in every non-generalized condition. Essentially, it is a tactic of scheduling behavior change programs in every condition to which generalization is desired; or at least to those on which baselines have been measured.

(3) Introduce to natural maintaining contingencies. Briefly, this is selecting behaviors to train which can be expected to come under the influence of trusted, stable and natural contingencies in the subject's environment.

(4) Train sufficient exemplars. This is merely mastery of the exemplar taught, and the obvious route of teaching another, then another, then another, and so on.

(5) Train loosely. The literature in the field contains few examples of this type. Researchers have tried to maintain thorough control of stimuli; yet tight control of very restricted and repetitive handfuls of lists or formats may restrict generalization. There appears to be somewhat of a movement to diversify and expand exemplars.

I would refer you to the article (Stokes and Baer, 1977) for explanation of the others. The point is, some attention is being paid to the emergence of a technology of generalization; and we need to be aware of it in aphasiology.

Some additional "what-to-do" possibilities for enhancing generalization which appear somewhat more "clinical" in nature are presented in Table 4.

Table 4. Enhancing Generalization: "What-To-Do Possibilities"

1. Keep training more exemplars.
2. Diversify exemplars.
3. Make unclear the limits of training contingencies.
4. Use stimuli found in generalization settings.
5. Monitor generalization setting.
6. Develop self-recording, self-reporting.
7. Reinforce generalizations sometimes.

(Stokes and Baer, 1977)

Included among these suggestions are adding and diversifying exemplars (trained items); muddying the boundaries of training contingencies (it's OK to reinforce in the hall after a session); using stimuli found in the natural environment; and in general paying more attention to pragmatics and the reinforcers found in the natural environment.

Summary

In summary, I have attempted to present some information on multiple baseline designs and its variations across behaviors, across subjects and across settings. Obviously, single-case design cannot answer all clinical research questions. Some disadvantages of the various time-series strategies have been pointed out, and they must be considered carefully. In multiple baseline designs a basic assumption is that targeted behaviors are independent from one another. If they should happen to co-vary when treatment is initiated, the controlling effects of the treatment variable are subject to question, and all the limitations of A-B designs listed in Hersen and Barlow (1976) are important.

However, what appears to be a disadvantage can be turned to an advantage in clinical aphasiology. We cannot afford to make too many a priori assumptions about which behaviors are independent and which co-vary. By

the application of these strategies, we will find that out. We will begin to map the territory. We will discover the rules of co-variance and generalization (perhaps they will relate to psycholinguistic systems we already know something about), and discover which communication behaviors are truly independent. Then we can demonstrate that efficiency of relevant communication can be effectively manipulated when treatment is instituted. This can only have a positive and cumulative effect on the people we are charged with serving, and just as importantly, on those future clinicians destined to treat us.

Bibliography

Baer, D.M., Wolf, M.M., Risley, T.R. Some current dimensions of applied behavior analysis. JABA, 1, 91-97 (1968).

Barlow, D.H. and Hersen, M. Single-case experimental designs: Uses in applied clinical research. Arch. of General Psychiatry, 29, 319-325 (1973).

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

Elbert, M., Shelton, R.L., and Arndt, W. A task for evaluation of articulation change: one development of methodology. JSHR, 10, 281-288 (1967).

Epstein, L.H. and Hersen, M. A multiple baseline analysis of coverant control. J of Behav. Ther. and Experimental Psychiatry, 5, 7-12 (1974).

Gottman, J.M., McFall, R.M., and Barnett, J.T. Design and analysis of research using time series. Psychological Bulletin, 72, 299-306 (1969).

Hall, R.V., Cristler, C., Cranston, S.S., and Tucker, B. Teachers and parents as researchers using multiple baseline designs. JABA, 3, 247-255 (1970).

Hersen, M. and Barlow, D.H. Single Case Experimental Design. New York: Pergamon Press, 1973.

Hester, P. and Hendrickson, J. Training functional expressive language: the acquisition of generalization of five element syntactic responses. JABA, 10, 316 (1977).

Holland, A.L. and Sonderman, J.C. Effects of a program based on the Token Test for teaching comprehension skills to aphasics. JSHR, 17, 589-598 (1974).

Jones, R.R., Vaught, R.S., and Weinrott, M. Time-series analysis in operant research. JABA, 10, 151-166 (1977).

Kazdin, A.E. Methodological and assessment considerations in evaluating reinforcement programs in applied settings, JABA, 6, 517-531 (1973).

LaPointe, L.L. Base-10 programmed stimulation: task specification, scoring and plotting performances in aphasia therapy. JSHD, 42, 90-105 (1977).

Liberman, R.P. and Smith, V.A. A multiple baseline study of systematic desensitization in a patient with multiple phobias. Behavior Therapy, 3, 597-603 (1972).

Mohr, J.P., Sidman, M., Stoddard, L.T., Leicester, J., and Rosenberger, P.B. Evolution of the deficit in total aphasia. Neurology, 23, 1302-1312 (1973).

Rosenbek, J., Green, E., Flynn, M., Wertz, R.T., and Collins, J. Anomia: A clinical experiment in Clinical Aphasiology: Conference Proceedings 1977. R. Brookshire (Ed.), Minneapolis: BRK Press, 1977.

Sidman, M. The behavioral analysis of aphasia. J. Psychiat. Res., 8, 413-422 (1971).

Sidman, M., Stoddard, L.T., Mohr, J.P., and Leicester, J. Behavioral studies of aphasia: methods of investigation and analysis. Neuropsycholigia, 9, 119-140 (1971).

Stokes, T. and Baer, D. An implicit technology of generalization, JABA, 10, 349-362 (1977).

Wahler, R.G. Setting generality: Some specific and general effects of child behavior therapy. JABA, 2, 239-246 (1969).

Weigel-Crump, C. and Koenigsknecht, R. Tapping the lexical store of the adult aphasic: Analysis of improvement made in word retrieval skills. Cortex, 9, 410-418 (1973).

West, J.A. Auditory comprehension in aphasic adults: improvement through training. Arch. Phys. Med. and Rehab., 54, 78-86 (1973).