The Multiple Baseline Across Subjects Design: Proposed Use in Research

DAVID B. CENTER, PH.D.,* and ROBERT A. LEACH, D.C.‡

ABSTRACT

This paper examines the use of single-subject research methodology in clinical research. Specifically, the use of the multiple baseline design across subjects is discussed relative to its application to chiropractic research. Brief discussion of the history of the development of single-subject research methodology is presented along with discussions of its strengths and weaknesses, issues related to its use, and some of the fields in which it has been utilized. The multiple baseline design across subjects is described and explained and an illustrative example of its application to a field based chiropractic research problem is provided. Finally, the advantages of this methodology for a field such as chiropractic, largely dependent upon the scientist-practitioner model, are discussed. (J Manipulative Physiol Ther 1984; 7:231–236)

Key indexing terms: single-subject research methodology, clinical research, multiple baseline design, applied research, scientist-practitioner model, chiropractic research.

INTRODUCTION

One of the most common problems in field based research is obtaining sufficient control over all relevant variables and an adequate sample, in terms of size and randomness, to employ traditional experimental designs. As a result, little purely experimental research is produced by individuals involved largely in field-based, service delivery. However, it is these individuals who have the best understanding of the problems and issues that are in most need of research in applied disciplines.

One solution to the inappropriateness of field based settings for purely experimental research has been the development of what are called quasi-experimental research designs.[1] One particular type of quasi-experimental design, the time-series designs, has received a great deal of attention and development by service oriented, field based researchers. These designs have been adopted and extensively used in recent years by clinical psychologists and educators and to some extent by psychiatrists and physical therapists.[2–6]

The time-series designs, as adopted and modified by field based researchers, have come to be called intra-subject replication designs[7] or single-subject designs. This refers not to the use of a N of 1, but rather of the continuous measurement, treatment, and evaluation of each experimental subject individually. Briefly, the rationale for this approach is that the traditional group statistical approach to research requires large numbers of subjects who are placed in control and experimental groups, given a pre-treatment measure on the dependent variable, treatment applied to the experimental group, given a post-treatment measure on the dependent variable, and then the two groups are compared on the dependent variable. This evaluation is usually based on a statistical test for group differences. The single-subject researchers[8] and at least one chiropractic researcher[9] object to this approach on several grounds. First, large groups of subjects aren't always available and/or practical. Second, the use of a control group in clinical research may unethically necessitate witholding treatment from some subjects. Third, the failure to measure the dependent variable during treatment eliminates the possibility of initiating treatment modifications which might be indicated by continuous monitoring of the dependent variable. Fourth, measurement of the dependent variable at single points in time (pre and post-treatment measures) may provide a biased estimate of the level of the dependent variable if there is much variability in the dependent variable. Fifth, averaging the data on a group of subjects may obscure important individual effects. Finally, evaluation by a statistical criterion may be misleading by indicating a

* Associate professor of Special Education Mississippi State University
‡ Private practice of Chiropractic
Submit reprint requests to: Dr. Robert A. Leach, P.O. Box 1313, Starkville, MS 39759
Paper submitted July 23, 1984.

highly statistically significant result which is not clinically significant. There exists a rather large body of literature on the single-subject research approach[10-23] and the interested reader is urged to consult this literature for a more detailed discussion of the above points as well as other issues that are beyond the scope of this paper.

DISCUSSION

While there are a number of single subject designs used in clinical research,[8] the one that appears to hold the most promise for field based, chiropractic research is the multiple baseline design across subjects. In the following a brief description of this design will be presented, then a discussion of some of the important considerations in its use, and finally an example of its application to chiropractic research will be provided.

A multiple baseline design across subjects uses each subject as his own control. Thus, it is necessary to establish, for each subject, a baseline on the dependent variable. This requires continuous measurement, for each subject, of the dependent variable. Once a steady state baseline has been established for each subject the treatment is begun. However, the intervention begins for the first subject only, while the baseline phase continues for the other subjects. Once a clear treatment effect has been obtained for the first subject, treatment is begun for the second subject while the baseline phase continues for the other subjects. Once a clear treatment effect has been obtained for the second subject, treatment is begun for the third subject and so on across all the subjects (see Figure 1).

As you will note in Figure 1, a somewhat idealized illustration, a steady state baseline was established on all four subjects during the first three sessions and treatment was begun for the first subject at session four. Baseline continued for the other three subjects during sessions four through six. By session six a treatment effect was clearly demonstrated for subject one and treatment began in session seven for subject two. Baseline continued for the other two subjects during sessions seven through nine. By session nine a treatment effect was clearly demonstrated for subject two and treatment began in session ten for subject three. Baseline continued for the last subject during sessions ten through twelve. By session twelve a treatment effect was clearly demonstrated for subject three and treatment was begun for the fourth subject.

In applying this approach to research the first factor that must be addressed is the selection of an appropriate dependent variable. In general, the requirement for a dependent variable is that it must be a variable that can be operationally defined. This has generally meant that some variable that can be directly observed or measured

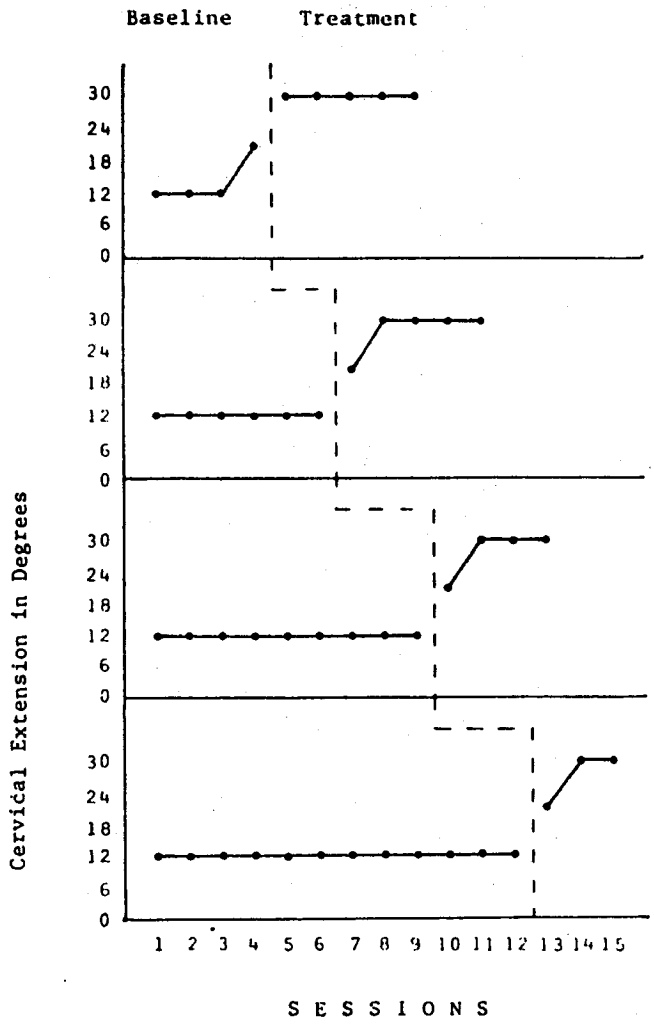

Figure 1. Introduction of treatment (independent variable) after first establishing a pretreatment baseline (dependent variable). Example: Improvement of cervical extension of osteoarthritic patients following twice weekly manipulation of the cervical spine.

with instrumentation is the best choice. For example, measures of range of motion or galvanic skin response (GSR) would be variables that could be objectively and repeatedly measured. Closely associated with the operational definition of the dependent variable is reliable measurement of the variable. Reliability of measurement requires measurement by two or more independent observers where measurement is by direct observation or careful calibration and monitoring of instrumentation. Reliability of measurement should be checked each time data is collected if possible but no less frequently than once per study phase.

The second consideration in a single-subject study is subject selection. Because this approach uses each subject as his own control the selection of subjects is a much simpler matter than the traditional experimental study. The main requirement is that each subject have

a problem that there is some reason to believe will be affected by the independent variable to be employed. Usually each subject has the same problem, although this is not mandatory. While there are some exceptions, (e.g., very rare disorders), the generally accepted minimum number of subjects required by the single-subject approach is three to five with the latter being preferred. This is considered a sufficient number for direct replication of the experiment and the establishment of limited generalizability.

The third consideration in a single subject design is a steady baseline for each subject. The baseline is established by repeated measures, over time, of the dependent variable prior to introduction of the independent variable. The minimum number of measures required for a baseline is three, since a trend in the data cannot be detected with any fewer. If there is much variability in the dependent variable more than three measures will be required. A suitable baseline should be relatively flat, show cyclical variability, be increasing (if the objective is to reduce), or be decreasing (if the objective is to increase). The baseline is a critical phase in a single subject study because it represents the data that the effectiveness of the intervention will be judged against. In fact, the baseline is being used to make a prediction about the future level of the dependent variable, and it is against this prediction that the independent variable is evaluated.

Thus, the longer and more stable the baseline the more confidence one can have in the prediction being made. Once the collection of baseline data begins it is necessary to graph the data. Graphed data is employed because it is easier to monitor the level of the dependent variable through a visual representation and because evaluation of the data will, in part, be done through graphic analysis. When constructing a working graph, the rule of thumb is that the length of the abscissa should not extend beyond two-thirds the length of the ordinate. A formal presentation of graphed data requires a squared graph as well as adherence to other graphing conventions. (33)

A fourth consideration in a single subject study is the evaluation of the data. Evaluation is accomplished using several criteria. First, the graphed data is examined for discontinuous phases, i.e., does the level of the dependent variable, in the treatment phase, differ markedly from the predicted level based on the data in the baseline phase. This decision is usually made using what is sometimes referred to as the "eye-ball" test for significance, i.e., a judgment is made based on visual inspection of the data. There has been considerable debate recently over the need to use statistical criteria in making this judgement, and the reader is referred to the literature already cited. Second, a treatment objective is usually set prior to beginning the study to establish a criterion for successful treatment[17,24] against which the effect of the independent variable can be judged. Third, the effect of the independent variable is judged against a social criterion, i.e., consumer satisfaction.[25,26]

A final point that needs to be discussed is the issue of internal and external validity. This will be only briefly discussed, but a more detailed discussion can be found in the literature cited.[7,8,21] The major threat to internal validity that will be discussed is "history". The most important thing that an experimental design must do is control for the possibility that some extraneous variable is responsible for the observed effect in a study. The multiple baseline design across subjects controls for "history" by using baselines on several subjects and introducing the independent variable at different points in time for each subject. Thus, to explain the effect produced by the independent variable by an extraneous variable requires separate extraneous variables affecting each baseline in the same way at different points in time. The probability of this occurring is not great, and it is therefore more parsimonious to attribute the observed effect to the independent variable.

The major issue relative to external validity to be discussed is generalizability of the results. As mentioned earlier, the requirement for three to five subjects to obtain direct replication establishes some limited generalizability. However, the single subject approach relies largely upon systematic replication for establishing generalizability. Systematic replication requires a body of literature to be established in which the effect of an independent variable is demonstrated across a large number of heterogenous subjects, a variety of service providers, and different presenting problems or variations on a presenting problem.

The following represents an example of how a study might be set up using the multiple baseline design across subjects in a piece of clinical research on chiropractic manipulative therapy.

EXAMPLE OF CHIROPRACTIC STUDY

Introduction

A common presenting complaint dealt with by chiropractic manipulative therapy is the lumbar disc lesion (annulus injury, nuclear bulge, nuclear protrusion, nuclear prolapse). Patients presenting with this disorder typically have either a history of acute trauma or insidious onset. Low back and hip pain is present which generally radiates into one leg. Physical examination of such patients usually reveals singular or multiple areas of muscle spasm with varying degrees of rigidity. There are range of motion limitations, paresthesias radiating

dermatomally, motor and/or sensory abnormalities, loss of lumbar lordosis visualized radiographically, and positive neurological findings.[27] Prognosis is dependent upon the age of the patient, prior history, physical development and overall state of health, and complicating factors such as prior lumbar surgery. In acute cases this disorder is sometimes self limiting; however, surgical intervention with some degree of permanent disability is often the outcome.[28] The purpose of this study is to evaluate the effect of chiropractic manipulative therapy (in this study mobilizing traction and cryotherapy is the treatment of choice) for reduction of the lumbar disc lesion. The experimental hypothesis is that mobilizing traction of the spine with cryotherapy and bracing will correct lumbar disc lesions more efficiently than conservative measures without mobilizing traction.

Methods

Ten subjects will be selected from the patient population who have presented the complaint described above and who meet the following criteria. Subjects will be screened for evidence of organic pathology, uncomplicated sprains/strains, spinal stenosis, degenerative joint disease with foraminal encroachment, and facet syndrome. In addition, subjects with symptomatology of less than one month's duration or who have been recovering spontaneously will be eliminated from consideration. Finally, subjects with history of prior lumbar surgery will be eliminated from consideration. Essentially pain or paresthesias or numbness has to have a dermatomal distribution and be localized to one specific dermatome.

Diagnosis of the disc lesion will be after Cox,[28] and only patients with three of the five following categories will be evaluated in this study:

1. Dominant symptom of leg pain of dermatomal distribution.
2. Paresthesias localized to a dermatomal distribution.
3. Straight leg raising reduced to 50 percent of normal, or pain crosses over to the opposite leg when the asymptomatic leg is elevated, or both. Also, if proximally or distally radiating pain is produced by digital pressure on the tibial nerve in the popliteal fossa.
4. Two of four neurologic signs present (i.e., muscle weakness, wasting, altered or diminished reflex activity, diminished sensorial perception).
5. Positive contrast study that corresponds to clinical findings.

Measurement

Range of motion will be defined as the degree of flexibility of the lumbar spine. This will be assessed using the standardized criteria adopted by the American Medical Association as set forth in the Guides to the Evaluation of Permanent Impairment. Standard goniometric technique will be utilized.

Nerve root embarassment will be defined as sensory or motor involvement of the spinal nerve root. This will be measured two ways. Vertical straight leg raising will be assessed using a standard goniometer for specific evaluation in terms of degrees of motion before pain is experienced, as well as before resistance is experienced. Differential deep tendon reflexes in the subject's legs will be assessed using a standard goniometer for evaluation. This technique will be used in order to objectify the measurement.

Reliability of the measures will be assessed by having two trained observers independently record the measures obtained on each of the three measures. Prior to beginning the study the assessor and reliability recorder will be trained using simulated observations. A minimum interobserver reliability of .80 for three consecutive sessions on the three measures will be required before training and modification of the measurement procedures will be considered complete. Interobserver reliability will be computed using the scored interval computation technique.

In addition to the measures described above a social validation evaluation will be done. This is essentially an evaluation of patient satisfaction. This assessment will be accomplished using seven point, bi-polar Likert type rating scales completed by the subject. These scales will cover such variables as experienced pain under varied circumstances, perception of sensory changes, and improvements in the physical dimension of the subject's life-style, i.e., work, recreation, etc.

Design

The multiple baseline across subjects design discussed earlier in this paper will be used. Following selection of ten subjects the study will begin. During the baseline phase each subject will be assessed during each session on the three measures previously described. In addition, each subject will complete the social validity scales during baseline. During this phase a placebo treatment would normally be introduced to control for expectation effects. However, for ethical reasons in this study conservative measures would be utilized including in home use of cryotherapy and bracing. Once a steady state baseline has been obtained on the subjects the independent variable will be introduced for one of the subjects. Measurement of the dependent variables will continue on a session by session basis. The remaining subjects will continue to receive conservative therapy and will continue with session by session measurement of the dependent variables. As soon as definite improve-

ment in the subject receiving the independent variable is demonstrated, the independent variable will be introduced for the second subject. The remaining subjects will continue in the baseline condition and will be systematically introduced to the intervention condition following the procedure just described.

Criteria for termination of treatment will be established prior to treatment based upon the extent of impairment evident in each of the subjects, relative to the dependent variables. After each subject is placed in the intervention phase, treatment and measurement will continue until the termination criteria are reached or it is clear that the treatment is not producing a positive effect. Should the latter occur, the subject will be returned to the baseline condition and once a steady state baseline has been demonstrated, a modification in the treatment or a new treatment procedure will be introduced, or referral made.

Data Analysis:

Data analysis will be accomplished by using both graphic analysis and statistical analysis. A squared graph will be constructed on each subject representing the session by session measurement of the three dependent variables. These graphs will be monitored on a continuous basis for ongoing evaluation of the dependent variables. Evaluation of the independent variable will be done by visual inspection for markedly discontinuous phases between the baseline phase and the intervention phase for each subject. Due to the small size of the study sample a nonparametric statistic, the Walsh test will be used as a final test for a significant difference between the data obtained in the baseline phase and in the intervention phase.

Because standard conservative therapy (i.e., cryotherapy and bracing) will be used during the baseline phase as a contrast for the chiropractic mobilizing traction in the intervention phase, the requirement for a steady state baseline may not be met. In other words, improvement in the subjects' conditions may be evident during the baseline phase. If this should occur, interpretation of the data could be somewhat confounded. The data analysis under this condition, will also include a trend analysis to determine if there was a significant increase in the rate of improvement after the experimental variable was introduced.

Procedures

During the baseline condition the subjects will receive standard conservative therapy which would include in home use of cyrotherapy and supportive bracing.

During the intervention condition the subjects will receive chiropractic traction therapy. Utilizing the "Cox"[28,29] method of flexion-traction on a standard flexion-traction therapy table, after appropriate cryotherapy (five minutes), the patient will receive ten minutes of distraction. Treatment will be augmented by continued use of in home cryotherapy and bracing.

Summary

The above example represents a possible application of single subject design to a chiropractic research problem. Specifically, the example shows how a multiple baseline design across subjects could be used to determine the effect of chiropractic mobilizing traction (flexion distraction) on the disc lesion with only ten subjects. Considering the type of research problems facing the chiropractic profession, such a design could prove invaluable.[30,32]

CONCLUSION

The single subject approach to research is one that has several distinct advantages for field based research.[8] First, the research effort is decentralized and allows the creative efforts of a larger number of researchers to be brought to bear on problems and issues within a discipline. This decentralization of research makes it possible for individual research projects to be carried out on much smaller budgets, although there may be no savings relative to the overall research effort. Second, the approach allows for a much needed merger between the roles of scientist and practitoners. Third, the pragmatic needs of the practicing clinician are more easily met through single-subject research. Finally, the professional's need for publication and recognition is more easily met through this approach by allowing for a larger number of "players".

Thus, the authors believe that single-subject research methodology and the multiple baseline across subjects design in particular represents a potentially valuable methodology for clinical research in chiropractic practice. Interested readers are urged to study the large body of existing literature on this methodology as it has been applied to research in other disciplines and to apply it to the research problems and opportunities available through their own practices.

REFERENCES

1. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research and teaching. In: Gage NL., ed Handbook of research on teaching. Chicago: Rand McNally, 1963.
2. Barlow DH, Hersen M. Single-case experimental designs: uses in applied clinical research. Arch Gen Psychiatry 1973; 29:319–25.
3. Gurainick M. The application of single-subject research designs to the field of learning disabilities. J Learning Dis 1978; 11:24–30.
4. Leitenberg H. The uses of single case methodology in psychotherapy research. J Abnorm Psychol 1973; 82:87–101.
5. Martin J, Epstein L. Evaluating treatment effectiveness in cere-

bral palsy: single-subject designs. Phys Ther 1976; 56:285–294.
6. Tawney J, Gast D. Single-subject research in special education. Columbus, Ohio: Merrill Pub, 1984.
7. Sidman M. Tactics of scientific research. New York: Knopf, Inc. 1960.
8. Hersen M, Barlow D. Single case experimental designs. New York: Pergamon Press 1976.
9. Vitelli M. A view on the fundamentals of chiropractic research methodology. Todays Chiro 1979; 8:9–11.
10. Azrin NH. A strategy for applied research: learning based but outcome oriented. Am Psychol 1977; 32:140–49.
11. Baer DM. Perhaps it would be better not to know everything. J Appl Behav Anal 1977; 10:167–72.
12. Baer DM. On the relation between basic and applied research. In: Catania AC, Brigham TA, eds. Handbook of applied behavior analysis. New York: Irvington Publishers, Inc. 1978.
13. Deitz SM. Current status of applied behavior analysis: science versus technology. Am Psychol 1978; 9:805–14.
14. Edington E. Nonparametric tests for single-subject multiple schedule experiments. Behav Assess 1982; 4:83–91.
15. Hawkins RP, Doles RW. Behavioral definitions in applied behavior analysis: explicit or implicit? In: Etzel BC, LeBlanc JM, Baer DM, Eds. New developments in behavioral research: theory, method and application. Hillsdale New Jersey: Lawrence Erlbaum Associates, Publ. 1977.
16. Kazdin A, Kopel S. On resolving ambiguities of the multiple baseline design: problems and recommendations. Behav Ther 1975; 6:601–08.
17. Kazdin A. Methodology of applied behavior analysis. In: Brigham T, Catania AC, eds. Handbook of applied behavior research; social and instructional processes. New York: Irvington Publishers, Inc. 1978.
18. Kazdin A. History of behavior modification: experimental foundations of contemporary research. Baltimore, MD: University Park Press 1978.
19. Kazdin A. Methodological and interpretative problems of single-case experimental designs. J Consult Clin Psychol 1978; 4:629–42.
20. Kazdin A. Research design in clinical psychology. New York: Harper and Row 1980.
21. Kratochwill T. Single subject research: strategies for evaluating change. New York: Academic Press 1978.
22. Hugdahl K, Ost LG. On the difference between statistical and clinical significance. Behav Assess 1981; 3:289–295.
23. Schwarzmueller E. Nonparametric statistics for behavior modification data (unpublished thesis). Atlanta, GA: Georgia State University 1974.
24. Rosenfield S, Houtz J. Evaluation of behavior modification studies using criterion referenced measurement principles. Psychol Record 1976; 26:269–278.
25. Kazdin A. Assessing the clinical or applied importance of behavior change through social validation. Behav Mod 1977; 1:427–52.
26. Wolf MM. Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart. Invited address given to the Division of Experimental Analysis of Behavior at the meeting of the American Psychological Association. Washington D.C. 1976.
27. Haldeman S. Spinal manipulative therapy in the management of low back pain. In: Finneson BE, ed. Low back pain. Philadelphia: J.B. Lippincott Co., 1980.
28. Cox JM. Low back pain: recent statistics and data on its mechanism diagnosis and treatment from chiropractic manipulation. ACA J Chiro 1979; 13:5125–5141.
29. Cox JM, Shreiner S. Chriopractic manipulation in low back pain and sciatica: statistical data on the diagnosis, treatment and response for 576 consecutive cases. JMPT 1984; 7:1–12.
30. Leach RA. An evaluation of the effect of chiropractic manipulative therapy on hypolordosis of the cervical spine. JMPT 1983; 6:17–23.
31. Leach RA. The chiropractic theories—a synopsis of scientific research. Baltimore: Williams & Wilkins (in revision).
32. Leach RA. The chiropractic theories—discussion of some important considerations. ACA J Chiro 1981; 15:519–525.
33. Miller L. Principles of everyday analysis of behavior. Monterey, CA: Brooks/Cole Publ, 1980.