The Application of Statistical Analysis to Single-case Designs
A Discussion Session

Kurt P. Kitselman, Jon L. Deal and Robert T. Wertz
Veterans Administration Medical Center, Martinez, California

The statistical analysis of data obtained in single-case designs may be possible as well as desirable under certain circumstances, but there are a number of pitfalls and problems connected with the use of such approaches for investigating the efficacy of aphasia therapy. The purpose of this Round Table Discussion was to clarify some of the relevant issues and to determine some guidelines that we might use in applying statistical analyses to our single-case studies. We discussed the types of designs and analyses that may be used and the problems that can be encountered in interpreting results.

Types of Analysis. Analysis of variance (ANOVA) models have been employed to evaluate data obtained in single-case designs. However, the use of conventional statistical tests in such designs is inappropriate for two reasons. First, ANOVA models are not robust in handling the effects of serial dependency in time-series data. Second, evaluation of means ignores trends in the data and may, therefore, lead to incorrect conclusions. Another type of analysis employs a time-series analysis to evaluate trends in the data as well as changes in levels that occur across treatment phases. Time-series analysis takes into account serial dependency. However, there are two problems with this approach. First, at least 50 to 100 observations must be obtained. Second, changes in the slope of the natural recovery curve complicate the analysis. One member of the group who used time-series analysis found it to be too cumbersome. Statistical analysis may also be used in the split plot method of trend analysis to evaluate changes in level and slope across treatment phases. Serial dependency is not a problem for this type of analysis. The only major problem pertains to patients who are still undergoing natural recovery. Changes in the slope of the natural recovery curve across time may complicate the interpretation of results and lead to incorrect conclusions in efficacy studies. The group decided that two designs have greater potential for usefulness in efficacy studies. The first randomizes treatment methods across sessions and obtains posttreatment test data after each session. The problem of serial dependency is ameliorated through the random assignment procedure. Thus, ANOVA models can be used to compare posttreatment test data obtained under different treatment conditions. Treatment effects must be at least partially reversible in order to show efficacy differences. The group was uncertain of the implications of this restriction for efficacy studies. Finally, the R_n statistic (Revusky, 1967) may be used to evaluate data obtained in multiple baseline designs. The serial dependency in time series data poses no problem for this analysis, and treatment effects need not be reversible. Of these applications of statistical analysis to single-case designs, members of the group had used two. The method that entails random assignment of treatments to sessions had been used with positive results. Time series analysis had been used and found to be excessively cumbersome.

Why Statistical Analysis? The first question we discussed was what advantages the use of single-case designs with statistical analysis might hold over the use of more conventional single-case designs. The only real advantage is that we may show treatment effects or differences that are too small to detect with conventional single-case designs. Identifying the main advantage led to another question: Are we interested in small effects? The consensus was that the outcome of treatment must be "clinically significant" in order to be of interest. While there was no agreement on what "clinically significant" means, there was agreement that the definition should relate to "functional communication." The problem of selecting appropriate outcome measures is relevant not only to efficacy studies using single-case designs with statistical analysis but to conventional withdrawal and reversal designs and to group studies as well. Selection of outcome measures may be less of an issue with conventional single-case designs because we have no opportunity to show small treatment effects. We do have this opportunity when applying statistical analyses to single-case designs, so we must ask if we are interested in small effects. Kazdin (1976) may help in answering this question. He wrote that the issue of clinical versus statistical significance is irrelevant when the treatments under comparison entail equal time and effort. In studies of treatment efficiacy, we hope to see clinically significant effects when comparing treatment with no treatment, i.e., when determining whether there is a treatment effect. We may be satisfied with statistical significance alone, however, when comparing treatments, i.e., when determining whether two treatments differ in their effects. The group could not agree. If statistically significant differences were found only in the testing of treatment items, this would not be satisfactory for some. For others, it would.

More on the Different Levels of Outcome Measures. The problem of defining appropriate outcome measures in efficacy studies was of central concern to the group. This issue pertains to group studies as well as to single-case designs, but it is of special significance in the latter because the small number of observations obtained in single-case designs may result in an unfavorable probability of showing treatment effects or differences that do exist when using certain levels of measurement. Possible measures suggested for use in single-case designs spanned a broad continuum of similarity between the outcome measure and the therapy. The measure most likely to show a difference between two treatment methods comes from one end of the continuum. This measure entails evaluation of performance during treatment sessions. However, measures obtained during treatment sessions probably cannot reflect on the relative efficacy of two treatment methods. The method that results in better performance during treatment sessions is not necessarily the more efficacious. The next step on the continuum of measurement levels can provide information on relative efficacy. At that step, items that were treated are tested after treatment sessions. The group thought it critical that outcome measures indicate the degree of generalization that takes place. A third level of measurement accomplishes this by testing items similar to those that were treated. At the far end of the continuum is some measure of "functional communication." It was agreed that comparing measures from this end of the continuum will reduce the chances of showing treatment effect differences that do exist. It was concluded that scores obtained during treatment sessions cannot provide information on the relative efficacies of different treatments. The measure

capable of determining relative efficacy must come from at least the next level. That is, outcome data must be obtained in posttreatment testing or observations.

Summary. The group discussion led to some conclusions regarding the types of analysis and design that may be useful in efficacy studies. ANOVA models cannot be used with conventional withdrawal and reversal designs because the data obtained in these designs violate critical assumptions of the ANOVA models. Time series analysis requires more observations than we can usually obtain, and the split plot method of trend analysis poses problems when studying patients who are still undergoing natural recovery because we do not know how to predict changes in the slope of the natural recovery curve. Randomization tests and the R_n statistic for evaluating data obtained in multiple baseline designs appear to have the best potential for use in efficacy studies. We also agreed that measures obtained only in posttreatment testing or observation can answer questions of relative efficacy. There was no consensus, however, on the types of measures that can provide convincing data on the efficacy of treatment or on the efficacy of one treatment compared with another. This is a problem we need to resolve for all efficacy studies, not just in single-case designs.

REFERENCES

Kazdin, A.E. Statistical analysis for single-case experimental designs. In M. Herson and D.H. Barlow (Eds.), Single-case Experimental Designs: Strategies for Studying Behavior Change. New York: Pergamon Press, 1976.

Revusky, S.H. Some statistical treatments compatible with individual organism methodology. Journal of the Experimental Analysis of Behavior, 10, 319-330, 1967.