Group Designs for the Study of Aphasia

Robert T. Wertz
Veterans Administration Hospital
Martinez, California

Jay Rosenbek
Veterans Administration Hospital
Madison, Wisconsin

At times, the study of aphasia appears to have been described, quite adequately, by one of our contemporary philosphers,

> We don't know where we are going or how we'll get there, but we know one thing--when we get there we'll be there. And that's something, even if it's nothing. (Perelman, 1947)

Two of the possible ways of "getting there" are to study aphasia with a group design or with a single case design. That is what this session is about, "Single Case and Group Designs for the Study of Aphasia," and the task of the first paper in the session, this one, is to discuss "Group Designs for the Study of Aphasia." The task will be accomplished by defining single case and group designs, discussing the differences between the two, elaborating the advantages and disadvantages of group designs, listing variations in the group designs, and speculating about combining designs to study aphasia. Some may call this approach naive or simplistic. We prefer to look upon it as basic or solid. We are on our way to getting "there."

Definitions

The group design is also known as the typical subject design. It permits the average (mean, median, or mode) performance of a group of subjects subjected to an experimental condition or conditions to be determined reliably. Silverman (1977) suggests that the minimum number of subjects necessary for a group design is approximately 10. Statisticians, especially biostatisticians, have a method for estimating the sample size necessary to achieve a given power. It requires the investigator to state how large a clinical difference must be observed to be considered significant (e.g., 10 percentile units in the PICA Overall score); the acceptable risk of not showing this difference; and how much error variance is expected in the measure used. These data allow one to compute the necessary sample size. Typically, this exercise is not done, since two main ingredients--the significant clinical difference and the error variance for a given measure with a given population--are unknown. Thus, "10 or more" is a popular estimate of the number of subjects necessary to conduct a group study.

The single case design is also known as the individual subject design or the single subject design. It permits the performance of an individual subject subjected to an experimental condition or conditions to be determined reliably. The necessary sample size is one. Little argument exists regarding how sample size is estimated.

Differences Between Designs

At least eight differences between single case and group designs exist. These are listed in Table 1. Single case studies provide data on the "typical" behavior of a single subject, and group studies provide data on the behavior of a "typical" group member. In the single case study, it is not necessary to assume that subjects respond similarly to an experimental condition. In the group study, it is necessary to assume that subjects respond similarly to an experimental condition. Single case studies require that subjects be run more than once under each experimental condition. Group studies require that subjects be run only once under each experimental condition. A single case design permits generalization to the "typical" behavior of the individual studied. Group studies permit generalization to the "typical" behavior of the mean or median group member studied. Sequence and order effects are difficult, if not impossible, to control in a single case design. In a group design, these are relatively easy to control. Few statistical procedures exist to assess the reliability of results obtained in a single case design. Many statistical procedures exist to assess the reliability of results obtained in a group design. Required sample size differs between designs; one in the single case design, and ten or more in the group design. And, finally, a single case design permits generalization to the population on a logical basis, while a group design permits generalization to the population on a statistical basis. These differences, as they pertain to single case designs, will be elaborated by the other papers in this session. We will concentrate on the differences that typify group designs.

TABLE 1
DIFFERENCES BETWEEN SINGLE CASE AND GROUP DESIGNS

Single Case	Group
Provide data on the "typical" behavior of a single subject	Provide data on the behavior of a "typical" member of a group
Not necessary to assume subjects respond similarly to an experimental condition	Necessary to assume subjects in a group respond similarly to an experimental condition
Necessary for subjects to be run more than once under each experimental condition	Not necessary for subjects to be run more than once under each experimental condition
Can generalize to the "typical" behavior of the individual studied	Can generalize to the "typical" behavior of the mean or median group member studied
Difficult to control for order and sequence effects	Relatively easy to control for order and sequence effects
Statistical procedures for assessing the reliability of results are not well developed	Statistical procedures for assessing the reliability of results are well developed
Minimum number of subjects necessary is one	Minimum number of subjects necessary is approximately ten
Can generalize to the population from which the subjects are selected on a logical basis	Can generalize to the population from which the subjects are selected on a statistical basis

(Adapted from Silverman, 1977)

Advantages of Group Designs

Group designs have advantages. They permit determining the reliability of results, generalization to the population, controlling order and sequence effects, controlling extraneous variables, and inferring characteristics of the population.

Examples of each of these advantages can be found in the literature. For example, Deal and Darley's (1972) work on the influence of linguistic and situational variables on phonemic accuracy in apraxia of speech asked, among other questions, whether patients with apraxia of speech could predict their errors beyond random guess. Comparison of predicted errors and errors actually made was significant ($p. < .01$) indicating the apraxic patients studied were reliable in predicting their errors.

De Renzi and Vignolo's (1962) report on the Token Test contains data to demonstrate a group design's ability to generalize to the population. Nineteen of 19 patients who demonstrated no receptive deficits on traditional tests made errors on the Token Test, indicating the instrument's sensitivity for detecting receptive disturbances in aphasic patients. One can assume, from these results, that the Token Test can detect mild auditory comprehension deficits in the population of aphasic patients. This assumption has been strengthened subsequently by reports from others (Boller and Vignolo, 1966; Swisher and Sarno, 1969; Wertz, Keith, and Custer, 1971).

Group designs permit controlling for order and sequence effects. For example, Wertz and Porch (1970) investigated the influence of auditory stimulation on verbal output in aphasic patients. Since the design required the subjects to speak in two conditions, noise and quiet, it was necessary to control for order and sequence effects. In a group design, this can be done by assigning half of the sample to one order (e.g., noise followed by quiet) and the other half of the sample to a second order (e.g., quiet followed by noise).

Group designs permit controlling for extraneous variables that may influence results. In a two or more group study, this can be done by random assignment to groups. For example, in the Veterans Administration Cooperative Study on Aphasia (Wertz, Collins, Weiss, Brookshire, Friden, Kurtzke, and Pierce, 1978), patients were assigned randomly to either Group A, individual treatment, or Group B, group treatment. Statistical comparison of extraneous variables that may have influenced results--age, education, initial severity of aphasia--showed no significant differences ($p. < .05$) between groups at intake. It is comforting to know that randomization really works. A second approach for controlling extraneous variables in group designs is to "block" on those variables that may influence results. For example, if one is studying the effects of therapy and is concerned that type of aphasia may influence results, the design can "block" on type of aphasia by including an equal number of patients of each type in a treated group and a nontreated group. The influence of type, if any, can be retrieved later by employing appropriate statistical analyses.

Finally, group designs permit inferring characteristics of the population from which the sample was drawn. The Schuell, Jenkins, and Carroll (1962) factor analytic study of the Minnesota Test for Differential Diagnosis of Aphasia is an example of inferring characteristics of the population from group data. Their results have given us a definition to describe the population; "Aphasia is a general language deficit that crosses all language modalities and may or may not be complicated by other sequelae of brain damage."

Disadvantages of Group Designs

Group designs have disadvantages. First, one must assume that the members that comprise the group are similar, and this may be a fallacious assumption. Second, the availability of subjects may be limited. Third, the mean performance of the group may not be "typical" of the performance of individual group members. Fourth, selection bias may restrict generalization. Fifth, attrition may influence the results. And sixth, ethical considerations may negate the use of some group designs.

Examples of each of these disadvantages exist in the literature and in our experiences. One cannot always assume similarity among group members. If Kertesz and McCabe's (1977) 13 patients shown in Table 2, initially classified as demonstrating Wernicke's aphasia, were studied to determine the influence of a specific treatment technique on Wernicke's aphasia, the results would be questionable. Final classification of these patients indicates only seven remained within the original classification. Six patients changed over time, perhaps as a response to treatment, perhaps because of the natural evolution of their disorder, perhaps for other reasons. They were dissimilar, and their response to the treatment administered was probably dissimilar. A rose may be a rose, may be a rose; but Kertesz and McCabe's results indicate that a Wernicke's, may not be a Wernicke's, may not be a Wernicke's.

TABLE 2
LACK OF SIMILARITY AMONG GROUP MEMBERS DEMONSTRATED BY EVOLUTION OF APHASIC TYPES

INITIAL CLASSIFICATION	N	FINAL CLASSIFICATION	N
Wernicke's	13	Wernicke's	7
		Global	1
		Transcortical Sensory	1
		Anomic	4

(Adapted from Kertesz and McCabe, 1977)

One needs a group to do a group study, and the availability of subjects may negate the effort. We screened over 1,000 patients in the Veterans Administration Cooperative Study on Aphasia (Wertz et al., 1978) to find 67 who met our selection criteria. For a while, we thought we had eliminated the disorder by initiating the investigation.

A group design assumes that the mean performance of the group represents the "typical" performance of group members, and this may not be the case. Consider the data in Table 3 on a hypothetical study designed to evaluate the influence of treatment on apraxia of speech. Comparison of pre- and post-treatment group mean errors indicates a ten percent improvement. Examination of individual subject performance reveals that only one patient improved ten percent. In fact, only three subjects improved at all, and the rest remained the same.

TABLE 3

HYPOTHETICAL STUDY ON TREATMENT OF APRAXIA OF SPEECH DEMONSTRATING GROUP MEAN PERFORMANCE MAY NOT BE TYPICAL OF INDIVIDUAL GROUP MEMBERS

SUBJECT NUMBER	% ERROR PRE-TREATMENT	% ERROR POST-TREATMENT
1	30	30
2	100	40
3	40	40
4	10	10
5	30	30
6	90	60
7	20	20
8	20	20
9	40	40
10	20	10
Mean	40	30

Group designs permit generalization of results to the population, however selection bias may restrict generalization. For example, the selection criteria employed in the VA Cooperative Study (Wertz, et al., 1978) do not permit us to generalize to 85 year-old, nonveteran females demonstrating aphasia subsequent to trauma who receive three hours of programmed instruction beginning three months postonset. Generalization of results is limited to patients who would have met patient selection criteria.

In a single case design, if attrition occurs, the study is over. In a group design, attrition may influence the results. Consider the influence of attrition present in the hypothetical treatment study shown in Table 4. The treated group displays a 30 percentile increase in the mean PICA Overall score. The no treatment group shows no change. Before declaring the treatment efficacious, one must consider the influence of attrition. In the treatment group, the more severe members who were showing no improvement tended to drop out, and in the no treatment group, the less severe members who were improving tended to drop out. Thus, attrition probably influenced the results.

TABLE 4

HYPOTHETICAL TREATMENT OF APHASIA STUDY SHOWING INFLUENCE OF ATTRITION ON RESULTS

GROUPS	COMPLETERS $\bar{X}$ PICA OA %ILE		DROPOUTS $\bar{X}$ PICA OA %ILE	
	PRE-	POST-	PRE-	AT DROP
Treated	50	80	30	30
No Treatment	50	50	60	80

Finally, ethical considerations may exist in some group designs. This is

particularly true in group treatment studies where the design requires a no treatment group. Clinicians are loathe to withold treatment. This position, of course, is based on the assumption that treatment, in fact, works. Hersen and Barlow (1976) point out the faulty logic of this position, since if treatment were known to work, there would be little need to test it. One cannot find an aphasia treatment study that employed a randomly selected no treatment group. Since aphasia therapy exists, the prevailing opinion is that it must not be witheld.

Variations in Group Designs

Despite the disadvantages, group research on aphasia is conducted, and not all group designs are the same. There are at least two types. One is the group design employed to determine whether differences exist. The Warren, Hubbard, and Knox (1977) comparison of normal and aphasic patients on short-term memory search tasks is an example of this approach. A second is the group design employed to determine whether relationships exist. The Maly, Turnheim, Heiss, and Gloning (1977) correlation of brain perfusion data with neuropsychological test scores is an example of this approach.

Additional examples can be culled from the literature. The VA Cooperative Study looked at differences between two types of treatment for aphasia. While both treatment groups improved their PICA Overall percentile scores during the treatment trial, Group A, individual treatment, resulted in greater gains than Group B, group treatment. To demonstrate that we were not guilty of burying individual performance in group means (a problem with many group designs), we looked at the percent of patients in each group becoming worse, remaining the same, and demonstrating mild, moderate, and marked improvement. Group means tended to reflect what was happening to individual patients. The majority of patients, in both treatment groups, made moderate to marked improvement during the treatment trial.

Attempts to determine prognostic influences on recovery from aphasia is another example of the group design employed to establish relationships. The typical approach is to compute partial correlations among assumed prognostic indicants and the amount of improvement demonstrated. In the hypothetical example shown in Table 5, the negative relationship between age and recovery would indicate that younger aphasic patients improve more than older aphasic patients. The correlation between education and recovery indicates that the number of years spent in school has little influence on the amount of recovery obtained. The negative correlation between months postonset and recovery indicates more improvement occurs early postonset than later. And the positive correlation between amount of treatment and recovery indicates that more treatment results in more improvement. All of these examples have been taken from group designs reviewed by Darley (1972, 1975) that have been employed to determine whether relationships exist.

TABLE 5
HYPOTHETICAL EXAMPLE OF A GROUP DESIGN EMPLOYED TO DETERMINE RELATIONSHIPS BY COMPUTING PARTIAL CORRELATIONS AMONG PROGNOSTIC VARIABLES AND RECOVERY FROM APHASIA

VARIABLE	r
Age	-.50
Education	+.25
Initial Severity	+.50
Months Postonset	-.50
Amount of Treatment	+.50

Some designs both determine differences and establish relationships. Kertesz and McCabe's (1977) report is an example. Their data, summarized in Table 6, indicate that different types of aphasia have different prognoses. For example, the future is brighter for a patient with anomic aphasia than a patient with global aphasia. Further, their data indicate the relationship between recovery and a given type of aphasia. The relationship between anomic aphasia and excellent recovery is high and positive.

TABLE 6
EXAMPLE OF GROUP DESIGN THAT DETERMINES RELATIONSHIPS AND DIFFERENCES BY EXAMINING PROGNOSIS FOR DIFFERENT TYPES OF APHASIA

APHASIA TYPE	N	RECOVERY (NUMBER OF PATIENTS)			
		POOR	FAIR	GOOD	EXCELLENT
Global	12	10	2		
Broca	12		5	3	4
Conduction	4			1	3
Wernicke	7	3	1	2	1
Transcortical and Isolation	3		1		2
Anomic	9				9
TOTAL	47	13	9	6	19

(Adapted from Kertesz and McCabe, 1977)

Designs for Specific Questions

We began with a list of differences between single case and group designs. These differences imply that one type of design is not appropriate for answering all questions. Such an implication is probably correct. One needs to select the appropriate design (Table 7) for the question posed.

TABLE 7
THE APPROPRIATE DESIGN FOR SPECIFIC QUESTIONS

QUESTIONS	DESIGN	
	SINGLE CASE	GROUP
Technique Building	+	
Technique Testing		+
Systematic Replication	+	
Actuarial Questions		+
Modification of Individual Behavior	+	
Modification of Group Behavior		+
Comparison of Two Treatment Packages		+

For example, Leitenberg (1973) suggests that the single case design is a good way to begin a treatment study; however it may not be the best way to end it. The flexibility of single case research permits isolating individual sources of variability and determining the power of specific treatment techniques. Bergin and Strupp (1972) call this approach "technique building." Once the sources of variability have been identified and the power of the techniques established, a global treatment package can be tested with a group design. Thus, the individual case design appears most appropriate for "technique building," and a group design is the method of choice for "technique testing."

Group designs are expensive. If systematic replication of results is deemed necessary, the single case approach is desirable. In fact, the investigator would be wise to employ a systematic replication series with single subject designs before ever considering a group design. Such a series provides information about patients, settings, therapeutic variables, and other elements that may predict success in a group design. It avoids a premature group comparison design conducted under less than ideal conditions. Many of the "weak" results obtained in group studies may be explained by large inter-subject variability. This may be avoided if a series of single case replications is done to determine, for example, with whom a treatment may be effective.

Actuarial questions--those requesting information on the magnitude of an effect--require a group design. Just as an insurance company may want to know how often age predicts death rates, a society that funds aphasia research may demand to know how often treatment predicts improvement. If one can say 80 of 100 patients who received treatment improved and only 15 of 100 patients who did not receive treatment improved, we have a message that society can understand. A design of this type cannot say why a treatment works, but it can state the size of the effect obtained.

If one is concerned with modifying the behavior of an individual patient, a single case design is appropriate. Conversely, if one is concerned with modifying the behavior of a group, a group design is required. This is so simplistic it should not need stating. Unfortunately, it does. Many continue to apply results obtained from single case designs which are effective in modifying individual behavior to a group, and, just as inappropriately, many continue to take group results and apply them to the individual. We all have examples of the treatment technique that "cured" Mr. Boomis but has not worked with another patient since. And, we all have stacks of treatment "programs" that a group study reported were effective for treating aphasia, but do not work with any individual demonstrating the problem. If the question is, "What will improve auditory comprehension in the patient I see at 9:00 a.m. on Monday, Wednesday, and Friday?," then one should employ a single case design. If the question is "What factors keep aphasic patients from dropping out of treatment?" employ a group design.

A group design is desirable when one wishes to compare two treatment packages. In this case, each treatment package would include a relatively complex number of therapeutic variables, and each would differ in content, administration, and theoretical basis. It is not possible to compare the influence of traditional, stimulus-response manipulation of language deficits in individual therapy with the influence of group treatment designed to stimulate language deficits by employing a single case design. The appropriate approach is a between-group comparison.

Combined Designs

Single case and group designs are not necessarily incompatible. In fact, one can complement the other, and both can be employed to meet the challenges we face. For example, a recent Lancet editorial (1977) began with the position that assessing the value of aphasia therapy is virtually impossible, and, interestingly, ended with several suggestions for achieving the impossible. The suggestions offered--multicenter trials, limited questions, and gradually building a hard core of firm facts--permit combining single case and group designs to accomplish the impossible task of determining whether aphasia therapy is efficacious.

Having recently completed a multicenter trial with a group design, those of us involved have developed an appreciation of the single case design. Anyone who plans a multicenter, cooperative effort would be well advised to begin with a series of single case designs to build techniques, establish sources of variability, and determine reliability before initiating a multicenter effort. Once the techniques have been built, then they can be tested with a group effort. For example, a comparison of clinic and home treatment for aphasia could begin with a series of single case designs and, eventually result in a multicenter between-group comparison.

Limited questions also permit beginning with single case designs followed by a group study to control sequence and order effects. For example, the question addressed by Podraza and Darley (1977), "What are the effects of auditory prestimulation on naming in aphasia?" could be explored further by building several specific techniques of auditory prestimulation in single case designs and end with a group study that permits all prestimulation techniques to be presented to each subject under the controls for sequence and order possible in a group design.

Finally, group designs generated by single case research permit a systematic way of gradually building a hard core of firm facts suggested by the Lancet editorialist. Again, techniques are built in single case research followed by a group study to test generalization. The Rosenbek, Lemme, Ahern, Harris, and Wertz (1973) treatment for apraxia of speech in adults reported a series of essentially three single case designs. Deal and Florence (1978) added four more. Gradually, the technique is being developed, and, eventually, it should be ready for a test in a group study.

In Parting

The title of this session is derived from Holland's (1976) challenge, "Our patients deserve our data not our word." Single case and group designs for the study of aphasia are ways of providing the data requested. They differ. Both have advantages and disadvantages. The extent to which we use each appropriately will determine how solidly our data support our word.

REFERENCES

Bergin, A.E. and Strupp, H.H. Changing Frontiers In The Science of Psychotherapy. New York: Aldine-Atherton (1972).

Boller, F. and Vignolo, L.A. Latent sensory aphasia in hemisphere-damaged patients: An experimental study with the Token Test. Brain, 89, 815-830 (1966).

Darley, F.L. Treatment of acquired aphasia. In W.J. Friedlander (Ed.), Advances In Neurology Volume 7: Current Reviews of Higher Nervous System Dysfunction. New York: Raven Press, 111-145 (1975).

Darley, F.L. The efficacy of language rehabilitation in aphasia. J. Speech Hearing Dis., 37, 3-21 (1972).

Deal, J.L. and Darley, F.L. The influence of linguistic and situational variables on phonemic accuracy in apraxia of speech. J. Speech Hearing Res., 15, 639-653 (1972).

Deal, J.L. and Florence, C.L. Modification of the eight-step continuum for treatment of apraxia of speech in adults. J. Speech Hearing Dis., 43, 89-95 (1978).

De Renzi, E. and Vignolo, L.A. The Token Test: A sensitive test to detect receptive disturbances in aphasics. Brain, 85, 665-678 (1962).

Hersen, M. and Barlow, D.H. Single Case Experimental Designs. New York: Pergamon Press (1976).

Holland, A.L. The effectiveness of treatment in aphasia. In R.H. Brookshire (Ed.), Clinical Aphasiology Conference Proceedings: 1975. Minneapolis, Minn.: BRK Publishers (1975).

Kertesz, A. and McCabe, P. Recovery patterns and prognosis in aphasia. Brain, 100, 1-18 (1977).

Editorial. Prognosis in aphasia. Lancet, II, 24 (1977).

Leitenberg, H. The use of single-case methodology in psychotherapy research. J. Abnormal Psych., 82, 87-101 (1973).

Maly, J., Turnheim, M., Heiss, W.D., and Gloning, K. Brain perfusion and neuropsychological test scores: A correlation study in aphasics. Brain and Language, 4, 78-94 (1977).

Perelman, S.J. Westward, Ho! New York: Simon and Schuster (1947).

Podraza, B.L. and Darley, F.L. Effect of auditory prestimulation on naming in aphasia. J. Speech Hearing Res., 20, 669-683 (1977).

Rosenbek, J.C., Lemme, M.L., Ahern, M.B., Harris, E.H., and Wertz, R.T. A treatment for apraxia of speech in adults. J. Speech Hearing Dis., 38, 462-472 (1973).

Schuell, H.M., Jenkins, J.J., and Carroll, J.B. A factor analysis of the Minnesota test for differential diagnosis of aphasia. J. Speech Hearing Res., 5, 349-369 (1962).

Silverman, F.H. Research Design In Speech Pathology and Audiology. Englewood Cliffs, N.J.: Prentice-Hall, Inc. (1977).

Swisher, L.P. and Sarno, M.T. Token test scores of three matched patient groups: Left brain-damaged with aphasia; right brain-damaged without aphasia; non-brain-damaged. Cortex, 5, 269-273 (1969).

Warren, R.L., Hubbard, D.J., and Knox, A.W. Short-term memory scan in normal individuals with aphasia. J. Speech Hearing Res., 20, 497-509 (1977).

Wertz, R.T., Collins, M.J., Weiss, D., Brookshire, R.H., Friden, T., Krutzke, J.F., and Pierce, J. Veterans Administration cooperative study on aphasia: Preliminary report on a comparison of individual and group treatment. A paper presented to the Annual Meeting of the American Association for the Advancement of Science, Washington, D.C., February, 1978.

Wertz, R.T., Keith, R.L., and Custer, D.D. Normal and aphasic behavior on a measure of auditory input and a measure of verbal output. A paper presented to the American Speech and Hearing Association Convention, Chicago, Ill., November, 1971.

Wertz, R.T. and Porch, B.E. Effects of masking noise on the verbal performance of adult aphasics. Cortex, 6, 399-409 (1970).