CHAPTER

1

Group Comparisons Across Neurologic Communication Disorders: Some Methodological Issues

Joseph R. Duffy
Penelope S. Myers

We would like to begin with a composite of some discussion that has or could have taken place within the last few years.

Question: On some of your measures, your bilateral people scored between the rights and the lefts. This makes me think it might be better to have bilateral than unilateral right- or left-hemisphere damage. This probably has something to do with different general severity among the groups. How do you go about matching severity if you think it's an important variable?
Answer: Well, I don't know how you do that.
Question: You found no differences among rights, lefts, and bilaterals, but you might have found differences if the groups were equated on severity. A lot of studies compare right to left to bilateral, but I don't think anybody has found a measure that you can really equate them on.
Answer: I agree.
Question: You found differences among your aphasic, right-hemisphere-damaged and closed-head-injured groups, but if the groups had been equated on severity, you might have found those differences.
Answer: Oh boy!

In 1978, considerable effort was devoted to methodological issues in the study of aphasia. Back then we seemed comfortable with group designs, or at least we used them more often than single subject designs. At that time, Wertz and Rosenbek (1978) reviewed the advantages and contributions of group designs and reminded us of their shortcomings. They focused primarily on studies of single groups of aphasic patients, comparisons among aphasic subgroups, and comparisons between aphasic and normal control groups.

Life wasn't so simple then, but now it seems like it was. In 1978 there were only faint stirrings of interest in nonaphasic neurogenic communication disorders. Today we accept the assertion that communication deficits can result from right-hemisphere dysfunction, and the increased prevalence of dementia and closed head injury has us studying their associated language and communication problems. As a result, these groups have gone beyond being useful only as controls in the study of aphasia. They have promoted an increasing awareness that numerous areas of cognitive function are involved in communication and have helped us learn that more than purely linguistic factors may explain aphasic behavior. They have become of interest in their own right.

Among other things, this means we are increasingly interested in making comparisons across groups of brain-injured people. The purpose of this chapter is to explore some issues associated with such comparisons. This seems worthwhile because group studies have the potential to improve our theoretical understanding of the cognitive and

neurologic underpinnings of human communication and its disorders, may improve differential diagnosis, and may help us to help those whose communication ability has gone awry.

SIMPLE GROUP COMPARISONS

We are all familiar with comparisons made between a single brain-injured group and a non-brain-injured control group. Such comparisons help to identify and describe the characteristics of a brain-injured group and may help test hypotheses about the nature of its deficits. For example, we can ask if an aphasic group is inferior to a control group on a dependent variable, such as verbal comprehension.

Once we have collected data on the dependent variable, we can learn

1. How each group performed.
2. Whether the two groups performed differently.
3. The incidence and range of abnormal performance in the aphasic group.

With some extra work, we can learn more. For example, with an independent measure of aphasia severity [e.g., an overall score on the Porch Index of Communicative Ability (PICA) (Porch, 1967)], we can establish the relationship between verbal comprehension impairment and aphasia severity.

If we add other dependent variables, such as tests of naming, repetition, and reading, we can describe a profile of performance for the control and aphasic groups. This profile can tell us if aphasic group performance simply represents a profile depression relative to the control group or if the shape of their profile is different from that of the control group. We also can determine if aphasic profiles are homogeneous or if subgroups exist, each with distinct profiles. This information can lead, and obviously often has, to subclassifications of aphasia.

PROBLEMS

External Validity

The first problem we have with these simple group comparisons is related to external validity. Brookshire (1983) has noted that the literature is remarkably inconsistent in subject description and selection. He iden-

tified 18 subject characteristics sometimes mentioned in published studies and argued that several of them (age, education, source of subjects, gender, lesion location, handedness, etiology, time post onset, severity of aphasia, and type of aphasia) should nearly always be reported. Without this information, Brookshire tells us, it may be difficult to generalize findings, resolve inconsistencies among similar studies, and replicate and extend valuable studies.

Test Sensitivity to Group Differences

Another potential problem relates to the sensitivity of tests to group differences. The ability to detect deficits is a function of the degree of deficit actually present as well as the sensitivity of the measure designed to detect the deficit. For example, it would be dangerous to draw broad conclusions about the presence of verbal comprehension deficits in aphasia using simple single-word comprehension tasks, measures on which a substantial percentage of aphasic patients would perform flawlessly. Other measures, such as the Token Test (DeRenzi and Vignolo, 1962), are much more sensitive to such deficits. We will return to this issue later because it is important for group comparisons that cut across different neuropathologies of communication.

The Meaning of Statistical Tests for Group Differences

Caution must be exercised in our interpretation of statistical differences. We seem to recognize the possibility that statistically significant differences may be clinically and theoretically meaningless. Some, perhaps many, statistical differences *are* clinically and theoretically meaningless, but we also should recognize that failure to find differences, or *apparently* meaningless differences, could be based on measures that are insensitive to the construct being examined.

Information about Uniqueness of Impairment

A major shortcoming of simple group studies is that they tell us nothing about uniqueness of impairment. For example, the only way we learned that the verbal comprehension deficits of aphasic individuals are tied to their aphasia, as opposed to the general effects of brain injury, was to establish that such deficits were not present or were present to a lesser degree or in a different form in groups with brain injury and no aphasia. Although a significant correlation between verbal comprehension deficit and aphasia severity can make us feel more confident about the deficit's relationship to aphasia, if we were just beginning to study the disorder,

we could argue that the relationship is a function of the general effects of brain damage as opposed to aphasia per se. Thus establishing that deficits are unique requires comparisons among brain-injured groups. Such comparisons lie at the heart of attempts to objectify differential diagnosis, as using similarities and differences among groups leads us to understand the contribution of various cognitive factors to intact and impaired communication.

CONTRASTING GROUP COMPARISONS

Comparisons across groups with presumed different neuropathologies of communication are not uncommon. For example, between 1978 and 1987, no fewer than 40 papers presented at the Clinical Aphasiology Conference have made comparisons across two or more brain-injured groups.

Ultimately, these studies help identify similarities and differences among groups, establish the discriminative power of various tests of communication ability, test theories about the nature of the communication deficits within and across groups, and, potentially at least, aid modifications in classifications of communication disorders.

PROBLEMS

There are several problems or challenges, it seems to us, associated with contrasting group comparisons. Some have obvious, if not always easy to achieve, solutions. Others are more perplexing but nonetheless worthy of discussion.

Classification Attributes

Let us assume that we are interested in comparing two or more of the following groups: aphasic, right-hemisphere-damaged, closed head-injured, and demented. The first problem is with the groups themselves, because the comparisons they invite are less analagous to apples versus oranges than to migrant workers versus B1 bombers. The basis for this is that they represent a mixing of classification attributes. Aphasia is defined by a constellation of behaviors, and membership in the group requires that set of behaviors. Membership in the right-hemisphere group is achieved by laterality of lesion. Membership in the de-

mented group requires a set of behaviors and often a progressive course, but without clearly defined or homogeneous locus or etiology. Finally, membership in the closed head-injured group is based on etiology.

The placement of patients into these different groups reflecting behaviors, lesion laterality, course, or etiology generates some problems for comparisons across the groups. That is, we cannot and perhaps should not match them on certain independent subject-description variables. For example, closed head-injured patients are typically young, and demented patients are usually older. Lesion locus can be described for the aphasic and right-hemisphere groups, but not for the demented without multi-infarct etiology and only partially or not at all for the closed-head injured. Etiology cannot be matched across all groups. Time post onset can be matched between right-hemisphere and aphasic groups with similar etiologies, can only be estimated in dementia, and may be difficult to interpret when comparing the recovery with plateauing associated with infarcts to deterioration in dementia and to the somewhat unpredictable and longer time course of recovery for the closed head-injured. Basically, these differing attributes produce a loss of experimental control and place restrictions on our ability to understand certain causal relationships.

Validity of Criteria for Group Classification

These group classifications also raise questions about the validity of the criteria used to separate the groups. Morris and Fletcher (1988), in an excellent discussion of problems of classification in neuropsychology, point out that any contrasting group study is a test of the validity of both the dependent and independent variables. In other words, when we compare groups on constructs relevant to a particular hypothesis and find differences, we usually conclude that the dependent variable is a meaningful dimension for group separation. Less obviously, it also can be inferred that the criteria used to separate the groups in the first place were valid.

A problem arises, however, when results are null, because there is no information in null results that pinpoints the basis for the absence of group differences. There may be problems in the original theory predicting group differences, problems with valid measurement of dependent variables, or problems associated with subject classification (invalid independent variables). Morris and Fletcher indicate that research tends to focus on the first and second problems but not the third. Why is debatable, but they note that many disagreements that seem theoretically motivated result, in part, from comparisons of overlapping or poorly defined groups. If nothing else, these observations suggest that group definition, description, and selection are as crucial to compari-

sons among groups of brain-injured patients as are hypotheses about dependent variables and their clinical assessment.

A second problem arises from the fact that the behavioral problems of each group are usually defined by different measures. For example, an aphasic group may be described by performance on a standard aphasia examination, a demented group by performance on a mental status examination, the closed head-injured group by ratings on a coma scale, and the right-hemisphere-injured patients by some measure of neglect. Thus these contrasting group comparisons add to Brookshire's list of subject description variables for studies of aphasia only. In addition, not only do these different measures measure different things, it is also not clear whether their sensitivity to each group's impairment is equivalent. We know that this problem of varying sensitivity exists across different tests of aphasia, but it is probably compounded significantly across tests of both independent and dependent variables measuring diverse behavioral disorders in different groups.

Equating Severity

To our knowledge, we have no standard metric for equating groups on some index of severity. The effect of this problem on the interpretation of research can be illustrated with the following example. Duffy, Duffy, and Pearson (1975) compared control, aphasic, and right-hemisphere-injured groups on measures of pantomime recognition, verbal recognition, and naming. The aphasic group was inferior to the right-hemisphere-injured and control groups on all measures, with no differences between the control and right-hemisphere-injured groups. This helped lead to the conclusion that pantomime recognition is impaired in aphasia and that aphasia may, therefore, be more than a linguistic deficit. Ignoring strong supporting correlation data for these conclusions, however, one could ask if the difference between the aphasic and right-hemisphere-injured groups had less to do with aphasia per se than with greater general severity of brain injury in the aphasic group.

Does this mean we are faced with having to derive an index of severity across which groups can be matched? If so, what is the construct? We could match aphasic and right-hemisphere-injured groups on size and site of lesion, but this may be unsatisfactory from a behavioral standpoint because some suggest (e.g., Semmes, 1968) that there are differences in neural organization between the two hemispheres. As Wertz, Dronkers, and Deal (1985) have suggested, one reason for difficulty in discriminating among brain-injured groups is that we have been using localization data to classify behavioral disorders. And size and site of lesion would not be helpful for demented or closed head-injured groups anyway.

Compounding Problems

These problems of group identification, measures used to describe groups, and equating groups on severity can each be compounded by additional problems in some group comparison studies. They include

1. Small N within groups.
2. Failure to establish test-retest reliability for dependent variables.
3. Failure to replicate, especially when N is small.

These problems are not unique, and we won't dwell on them. The solutions can be simple: Increase N, establish reliability of the dependent variables, and replicate.

SOLUTIONS

How can we minimize some of these problems? Obviously, the problems we've discussed aren't relevant to all studies, and some apparent solutions may even be contraindicated when certain questions are addressed. With these caveats in mind, let us review some possible solutions.

Improve Group Description

The problems of external validity in studies of aphasia alone can probably be multiplied in contrasting group comparisons by a factor equal to the number of groups being compared, especially because we are not always sure of the most relevant variables. Thus Brookshire's (1983) recommendations for subject description deserve special attention. The more information we have about each group's characteristics, the better we can judge how well each group represents the population from which it was drawn. At the least, this can permit generalizations about the deficits exhibited within a group, even if contrasting group comparisons remain a problem.

In this context, we suggest that when tested abilities can vary as a function of premorbid intelligence, contrasting group comparisons should include an estimate of premorbid IQ for all subjects. Such estimates are now computable (Wilson et al., 1978) and may be important when dependent variables measure more than the very basic language abilities assessed by traditional aphasia tests.

Administer All Tests to All Groups

One wonders about the value of administering all tests used to describe each group to all groups. That is, should the PICA be given to a right-hemisphere-injured group in studies comparing aphasia with right-hemisphere-injured patients? Should both groups be given a test of neglect? Should all groups be given mental status examinations when group comparisons include demented patients? At the least, these numeric indices further describe each group and can serve as independent variables whose relationships to dependent variables also can be examined.

Don't Mix Etiologies within Groups

This applies to all groups but may be particularly important for demented groups and groups called "bilateral." Patients with multi-infarct dementia versus Alzheimer's disease may behave quite differently, and these two etiologies probably should not be combined in a single group. Groups defined as bilateral are, unfortunately, often characterized by several etiologies, and even when infarct is the cause, we often are not told if "bilateral" was the result of single or multiple events or whether the lesions are cortical or subcortical, supratentorial, or in the posterior fossa. The heterogeneity of these etiologic and localization factors could generate a lot more within-group than between-group variance and obscure or distort differences that might be detectable in a more narrowly defined sample.

Examine Relationships among Variables

Simply describing differences across groups on dependent variables again leaves open the problem of unequivalent groups. Use of correlational and multiple-regression techniques to examine relationships among dependent and independent variables is a more powerful way of establishing the ability of various measures to distinguish among groups and to examine the causal relationships among variables presumed to measure different and similar underlying cognitive abilities.

Examine Error Types

Our efforts at numeric objectivity may obscure differences among groups. When the purpose of an across-group study is to examine differences among groups, it may be very useful to examine error types. Tests can be failed miserably for very different reasons, and as the PICA has demonstrated so ably, identifying the behavioral characteristics of

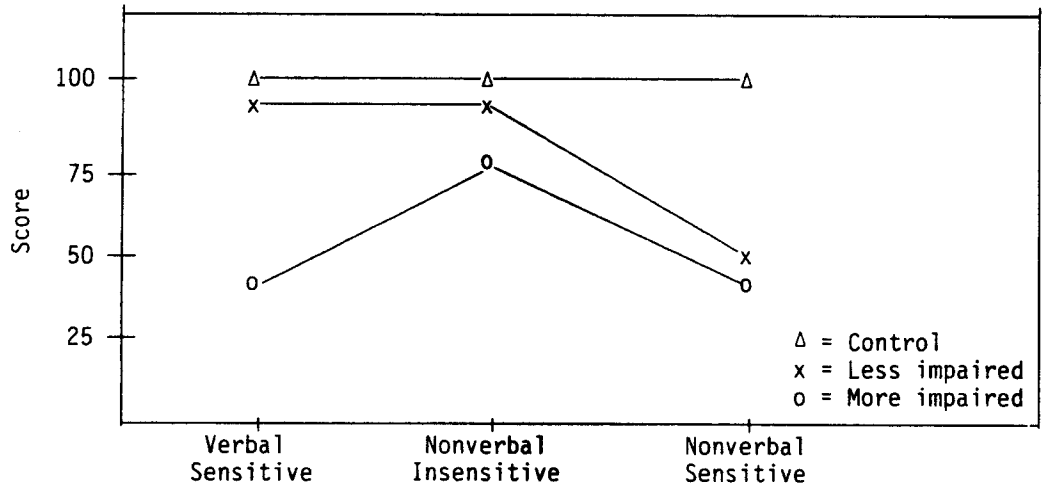
responses may capture a distinguishing feature not detectable by the simple outcome of a task. That is, there may be profiles within tasks and not just across tasks; to lose this information may be particularly damaging when comparisons are made across groups.

Equate Sensitivity among Dependent Variables

Another issue relates to group comparisons made across more than one dependent variable. In such instances, it may be important to establish equivalent sensitivity of the dependent variables, especially if the goal is to demonstrate that groups have different patterns of deficit. Brown et al. (1987) point out that erroneous conclusions may result if tasks are mismatched on properties known to affect their sensitivity to deficits. Of particular relevance, they also indicate that comparing sensitive to insensitive tasks can be especially troublesome when the general effects of brain injury are greater in one group than in another. The poorer performance of the more generally impaired group on the more sensitive task might be taken as evidence of a specific-unique impairment when it is not actually the case.

For example, Figure 1-1 illustrates what could happen if we compare two groups (0 and X) on a sensitive verbal task and an insensitive nonverbal task and group 0 is more generally impaired than group X. If group 0 is inferior to group X on the sensitive verbal task but not the insensitive nonverbal task, we might conclude that group 0 has a specific impairment of verbal abilities. However, if the verbal and nonverbal tasks were of equal sensitivity, we might discover that group 0 is just as

Fig. 1-1. Potential effects on comparisons between groups with different degrees of general impairment when tasks differ in sensitivity to specific impairment.



impaired nonverbally as verbally and, therefore, has an across-the-board impairment. To make things worse, we also might discover that the sensitive nonverbal measure detected a specific deficit in group X. Thus, with dependent variables of unequal sensitivity, we might have concluded that a significantly impaired group had a specific impairment when it actually had across-the-board deficits and that a less impaired group had no impairment when, in fact, it had a specific impairment.

Procedures for matching tests for sensitivity have been described by Chapman and Chapman (1973, 1978) and basically involve equating tasks on reliability, mean item difficulty, and shape of distribution of item difficulty. We cannot review the procedures here, but this issue should be addressed in studies seeking to establish differential patterns of performance across groups.

Test More Than One Hypothesis

Although not unique to contrasting group comparison studies, it also seems important to test more than one hypothesis. This gives overt recognition to the likelihood of multiple determinants of complex phenomena (Duffy, Watt, and Duffy, 1981) and, if done in a certain way, also may help us deal with the equating-of-severity issue. That is, hypothesizing an interaction or a kind of double dissociation between group and dependent variables, if the interaction is confirmed, may help alleviate concern about unequal severity of brain injury between groups. Returning to the Duffy, Duffy, and Pearson (1975) study, in which the aphasic group was consistently inferior to the right-hemisphere-injured group, this unequal-severity issue might have been addressed with the addition of a dependent variable *intended* to generate poorer performance by the right-hemisphere-injured group. That is, if the authors were able to get a dependent variable to reverse the differences between the aphasic and right-hemisphere-injured groups, they would have had additional evidence to suggest that a factor of general severity was not the entire explanation for the aphasic group's inferior performance on the dependent variables in which they were most interested. This strategy may be a powerful, and achievable, partial solution to the problem of equating groups.

FINAL ISSUES: EQUATING SEVERITY AND IMPROVING CLASSIFICATIONS

Do we need a behavioral measure on which groups can be equated or matched on severity? Is it possible to develop a measure or small set of

measures of basic cognitive skills which, when administered to a large number of individuals in each group, would generate similar means and distributions across groups? The identification or development of such a measure would provide a metric against which groups could be matched and compared and would help to strengthen conclusions drawn about the meaning of group differences and relationships on dependent variables under investigation. Recent literature suggests that this goal may be achievable because there are measures that have generated defective performance in more than one brain-injured group without generating differences between brain-injured groups (e.g., Bayles et al., 1989; Brown et al., 1987).

Finally, we return to the groups themselves and ask again about the problems they may be creating for us. *Aphasia,* to us, is the most satisfactory label we have been discussing, primarily because it defines the behavior of interest. The label *right-hemisphere-damaged* is particularly unsatisfactory; we suspect we have gone beyond the usefulness of this broad generic label. In fact, we wonder if it might not be time to generate a working label reflecting a theory of the communication deficits associated with right-hemisphere damage. This would require a definition that might include a constellation of behaviors whose presence would be required in studies of right-hemisphere-damaged patients who have communication deficits. At the least, we should be working toward a more narrowed definition of this right-hemisphere-damaged group, and we believe the definition should be based more on behavior than on neuroanatomy. One may raise the same issue about those with dementia and closed head injuries. Do we know enough about their communication deficits to define them by those deficits and then use the presumed disease or etiology simply as a descriptor?

The real issue here relates to theory and classification as much as to methodology. Our group divisions for the purposes of understanding the nature of communication disorders should be based on a theory about the deficits and their underlying cognitive bases. Morris and Fletcher (1988) summarize the enterprise well. They say that "the determination of pertinent variables for describing patients in various groups is an interactive process representing a search for theoretically meaningful measurement constructs, operations, and salient criteria for defining the various disorders of interest. Research and clinical assessment progress through the interaction of the development of theory, valid measurement tools, and valid classification systems" (p. 641). In other words, developing theory and improving differential diagnosis may work best when they work together.

SUMMARY

Across-group comparison studies are going to be with us as long as we are interested in the classification of communication disorders, their differential diagnosis, and understanding the basic nature of a variety of neurogenic communication deficits. We are not sure if all the things we have defined as problems are problems, and we are even less sure that our proposed solutions are the correct ones. We are sure that there are problems and solutions that we have not identified. Finally, we hope this discussion about apples and oranges has provided some food for thought, because we are also sure that what we learn from group comparison studies is only as good as the questions we ask and the methods we use to answer them.

REFERENCES

Bayles, K. A., Boone, D. R., Tomoeda, C. K., Slauson, T. J., and Kaszniak, A. W. (1989). Differentiating Alzheimer's patients from the normal elderly and stroke patients with aphasia. *Journal of Speech and Hearing Disorders, 54,* 74–87.

Brookshire, R. H. (1983). Subject description and generality of results in experiments with aphasic adults. *Journal of Speech and Hearing Disorders, 48,* 342–346.

Brown, G. G., Sawyer, J. D., Nathan, A., and Shatz, M. W. (1987). Effect of lateralized cerebral dysfunction on the continuous paired-associate test. *Journal of Clinical and Experimental Neuropsychology, 9,* 680–698.

Chapman, L. J., and Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin, 79,* 380–385.

Chapman, L. J., and Chapman, J. P. (1978). The measurement of differential deficit. *Journal of Psychiatric Research, 14,* 303–311.

DeRenzi, E., and Vignolo, L. A. (1962). The token test: A sensitive test to detect receptive disturbances in aphasics. *Brain, 85,* 665–678.

Duffy, J. R., Watt, J., and Duffy, R. J. (1981). Path analysis: A strategy for investigating multivariate causal relationships in communication disorders. *Journal of Speech and Hearing Research, 24,* 474–490.

Duffy, R. J., Duffy, J. R., and Pearson, K. T. (1975). Pantomime recognition in aphasics. *Journal of Speech and Hearing Research, 18,* 115–132.

Morris, R. D., and Fletcher, J. M. (1988). Classification in neuropsychology: A theoretical framework and research paradigm. *Journal of Clinical and Experimental Neuropsychology, 10,* 640–658.

Porch, B. E. (1967). *Porch index of communicative ability.* Palo Alto, CA: Consulting Psychologists Press.

Semmes, J. (1968). Hemispheric specialization: A possible clue to mechanism. *Neuropsychologia, 6,* 11–26.

Wertz, R. T., Dronkers, N. F., and Deal, J. L. (1985). Language and localization:

A comparison of left, right, and bilaterally brain-damaged patients. In R. H. Brookshire (Ed.), *Clinical Aphasiology,* Vol. 15 (pp. 141–148). Minneapolis, MN: BRK Publishers.

Wertz, R. T., and Rosenbek, J. (1978). Group designs for the study of aphasia. In R. H. Brookshire (Ed.), *Clinical Aphasiology,* Vol. 9 (pp. 1–10). Minneapolis, MN: BRK Publishers.

Wilson, R. S., Rosenbau, G., Brown, G., Rourke, D., Whitman, D., and Grisell, J. (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology, 46,* 1554–1555.