The Usefulness of Treatment for Aphasia: A Serendipitous Study

Audrey L. Holland
University of Pittsburgh, Pittsburgh, Pennsylvania

INTRODUCTION

Efficacy studies are hard to come by. The problems inherent in developing a thorough, adequately controlled study of the effectiveness of aphasia rehabilitation are great, possibly even insurmountable. Statistical matters are involved, such as adequate sample size and assuredness of random sampling procedures. Ethical questions abound, such as withholding treatment, particularly when the only investigators who seem to find the problem interesting enough to study also are the ones who believe that treatment helps. Control itself is a problem; controlling the host of factors such as age, aphasia-producing mechanism, spontaneous recovery, type, frequency, adequacy of treatment (to name a small few) also scare away all but the nerviest of investigators. Wertz (1980) has called attention to a 1977 Lancet editorial that discussed some of these problems in assessing treatment and concluded that until more is known about aphasia itself, investigations of treatment should concentrate on small, well defined studies comparing one mode with another. Darley's review (1972) of more than twenty such studies indeed suggests that treatment helps; Wertz' (1978) study, which is the most carefully controlled in the literature, clearly argues well for the usefulness of treatment.

But efficacy studies, after all, are not undertaken only for scientific, dispassionate motives. Efficacy studies are also undertaken to show, perhaps to ourselves, but also to referral and payment sources, that treatment has a substantial role in positively influencing recovery from aphasia, or the quality of an aphasic person's life. Basso, Capitani and Vignolo's study (1979) is quite unsatisfactory from a neuro-epidemiological point of view. Yet it apparently moved D. Frank Benson from his lofty position of withheld judgment and often thinly-veiled distaste, to declare that

> ...therapy does affect recovery from aphasia. How much of the improvement stems from psychic support offered by the therapy program and how much is due to actual language training techniques remains unknown. While the study may have technical deficiencies... it is so large, has sufficiently definite results and would be so difficult to improve upon that it commands respect. It would appear more fruitful to focus efforts on improving therapy techniques rather than on additional statistical refinements of the treatment-no treatment paradigm. (p.182).

All of us would, do, slather for our own version of the compelling study (of even a compelling single patient) to move our own versions of Frank Benson from the ever-so-boring question, "Does therapy help?" to a cry for treatment comparisons. But as La Pointe and the naive investigator he is fond of

quoting point out, "THEY (whoever they may be) have the right to demand our data instead of our word."

All of this is said to make a simple point. If an efficacy study creeps up on you, PUBLISH. The study I am sharing today is such an animal. It is entirely post hoc, stemming from an attempt to measure something else altogether, and therefore not victim to many biases or controls. It is offered in hope that it might be of use in convincing someone still stuck on the treatment-no treatment aspect of efficacy research.

METHOD

Two large research projects have recently resulted in the development of a new measure of functional communication for use with adults who have aphasia. The first project developed and validated the test, CADL (an acronym for Communicative Ability in Daily Living (1980)). In addition to the new test, aphasic patients who participated in the validating study received a battery of traditional aphasia tests. They were observed in their natural environments and interviewed concerning a number of matters related to physical condition, social interaction, and personal characteristics. Clinic medical records were also consulted. These data were used to validate the test. The second project revised the test and then developed norms on it related to age, type of aphasia, and whether or not a given patient was living at home or in an institutional setting--all of these features having been shown to be of interest in the original study.

Resulting from an item analysis done on the original, validating CADL, four items were deleted, and wording on some items changed. The present version, for which the norms are available, is thus shorter than the original by four items. The revision made it impossible to use the CADL data from the eighty aphasic subjects on whom the test was validated. However, twenty-eight of the original patients agreed to be tested again on the revised test. This second testing occurred roughly one year after the first; the shortest time elapsed between tests was eight months and the longest time, fifteen months. These second test scores were used in the norming study, which ultimately involved 130 aphasic patients.

There were 68 items common to both $CADL_1$ and $CADL_2$. These common items allowed us to analyze CADL's stability over this comparatively long time period, by correlating performance at times 1 and 2. Pearson product moment correlations between these two sets of scores resulted in a correlation coefficient of .96, statistically significant at .01. This finding was in line with CADL's previously determined test-retest reliability, which had been measured at a three-week interval for twenty chronic aphasic patients as part of the initial validity study.

While preparing these data for statistical analysis, it was noted that seventeen of the twenty-eight twice-tested patients had improved CADL scores, eight patients scored lower, and three remained exactly the same. Most of the changes were slight, as could be expected from the reliability data reported above. Nonetheless, the disproportionately high frequency of improved scores was intriguing. It caused us to investigate patient characteristics (all gathered at the time of the first test) that could explain it.

RESULTS AND DISCUSSION

A number of features were considered; these included age, whether or not the patients were institutionalized, months post-onset at time of test one, the type and severity of aphasia. For none of these variables did the improvers appear to differ from the larger group. For example, four of the sample's seven institutionalized patients improved; three did not. The BDAE severity range for the whole group ranged from 0-5; for the improved group, it ranged from 0-4. All types of aphasia seen in the whole group ranged from 4-72 months and for the improved group from 4-70 months. The oldest patient was among the improvers, the second oldest was not; the same was true of the youngest patients. One single variable, left handedness, occurred only in the improved group--three of the total group were left-handed and all improved. It was decided to follow through on this variable, by calling the clinics who treated the lefties to seek further information about them (more about this later).

The only thing of interest that we learned about the left-handers was that they had continued therapy for at least six months. We wondered if treatment could be a variable. We then called the six cooperating clinics who had furnished the whole sample of twenty-eight patients and asked them to check records and to inform us as to which of the patients they furnished had continued in treatment, like the lefties, for at least six months in the interim between their CADL tests. Fifteen patients had not continued in treatment for that time period; thirteen had. We then grouped aphasic subjects into continued and discontinued treatment groups, and lo!, twelve of the thirteen continued patients were among the improvers. We compared CADL change scores for continued and discontinued treatment groups using a t-test for independent samples. This analysis resulted in a statistically significant difference (.05) between the groups (Table 1).

Table 1. Mean CADL scores (computed in percentages--common items)

	Continued treatment group (N=13)	Discontinued treatment group (N=15)
time 1	70.46	70.40
time 2	77.31	68.93
mean difference	+6.96	-1.47

t-test for independent samples, computed on difference scores 2.93 (df=26) significant at .05

First, it must be emphasized that all the patients here represent chronic aphasias. All patients were at least four months post-onset at the time of the first testing. By the time of the second test, a minimum of eight months had elapsed, making the aphasic patients at least a year post-onset by then. Most of them were a great deal more chronic. Thus, spontaneous recovery can hardly be explanatory. In fact, these data support Sarno's recent comments (1979) regarding improvement in chronic, particularly

global, patients. She has suggested not only that they have slower recovery rates, but also that recovery extends farther in time for the severely involved patient.

Second, six clinical settings were involved. Four of the six furnished both continued and discontinued patients, and none of us expected to be scrutinizing these patients for gains following the first test. The normal clinical events followed the first test, and it is assumed they were based on good clinical judgment, not study considerations. This helps to eliminate bias in the study.

Third, the clinics included a variety of settings as well as approaches to treatment, so that the variable of type of treatment these patients received washes liberally through the data and cannot be used to explain the study's results. Treatment here--not a particular type of treatment--appeared to increase functional scores.

Finally, the continued patients include a variety of types of aphasia, including global patients. Even in such a small N study, lack of control of this factor did not overwhelm the treatment effect.

Let us turn now to the features that describe and to some extent differentiate the continued from the discontinued patients. Table 2 summarizes some of these, all measured at the time of the first test. The formal measures are roughly comparable, with the mean of the discontinued group being slightly lower. The discontinued patients are skewed toward more severe aphasias, they are slightly older, definitely more right-handed and longer post-onset of aphasia than are the treated group. More interesting, though, are the comparative ranges on continuously distributed variables. In all cases, a broader range typified the discontinued group. I believe these ranges reflect clinician reliance upon traditional criteria for continuation or dismissal from treatment. Patients who make more full recoveries or who are making no gains are discontinued, as are patients who have been treated for a very long time, or who are among a clinic population's oldest. The continued group, conversely, represents a subsegment of the aphasic population about whom test scores allow us traditionally to be more optimistic. It should also be noted that the continued ones are also those about whom we are not as eager to prognosticate: they are patients who are not as well known.

Table 2. Patient characteristics

	Continued treatment group (N=13)	Discontinued treatment group (N=15)
TYPE OF APHASIA		
Broca	5	4
Wernicke	3	4
anomic	3	1
mixed, global	2	6
HANDEDNESS		
right	10	15
left	3	0

Table 2. continued

	Continued treatment group (N=13)	Discontinued treatment group (N=15)
SEX		
female	2	5
male	11	10
AGE		
mean	55	60
range	32-70	41-84
EDUCATION (years)		
mean	11	11
range	6-16	8-20
MONTHS POST-ONSET (at first test)		
mean	15.23	27.40
range	4-30	4-72
BDAE SEVERITY RATING		
mean	2	2
range	1-4	0-5
BDAE AUDITORY COMPREHENSION		
mean	81	75
range	34-118	12-130
PICA (%ile)		
mean	61.92	58.00
range	25-93	10-96

Three further points--to be covered briefly. The first concerns the left-handed subjects who got me into this Monday-morning Terry Bradshawing in the first place. Two are anomic, one has Wernicke aphasia. Two are moderately-severely impaired, one is only mildly so. One of these aphasias (the mildest) is left-hemispheric stroke-induced; two have had left hemisphere surgery. None is hemiplegic. On tests, they look like their right-handed compatriots. Boller has recently suggested (in press) that even crossed aphasias do not differ in aphasic symptomatology very much from more typical patients. I would suggest the same for left-handers with left language dominance as well. Yet all three of them made it into the continued group, and gained from treatment as measured here. This supports the frequent contention that aphasia is less severe in left handers. That they were chronically aphasic denies the assertion that aphasia is more transitory among them.

The second point briefly to be covered is this. What about the patients, five of them, who continued to improve on functional testing without treatment? They also have little to unify them; they comprise three mixed (global), one Broca and one Wernicke aphasia; range in severity on the BDAE from 0-4;

and in months post-onset from 4-70 months. The oldest discontinued improver was eighty years old, an institutionalized man who was seventy months post-onset. In fact, the two institutionalized patients are the most interesting of the discontinued improvers--they have the longest months-post-onset time, have the lowest PICAs and are the oldest. One wonders whether traditional termination criteria were well applied in these two individual cases.

Finally, I have been making a lot out of improved functional scores. What was the nature of the functional change that was represented by improved scores? Given CADL's peculiarities, it is possible to look both at quantity and quality of responses, and it is of interest to summarize the item analysis along these two dimensions. By and large, both more information and more specific information was supplied by improvers on the second CADL test. Refinement of responding, as well as more of it, are typical of good treatment. And both are represented here.

REFERENCES

Basso, A., Capitani, E., and Vignolo, L., Influence of rehabilitation on language skills in aphasic patients. Archives of Neurology, 36, 1979, 190-196.

Benson, D.F., Aphasia, Alexia, and Agraphia. New York: Churchill Livingstone, 1979.

Boller, F., Crossed aphasia; A review of the literature and three new cases. Brain and Language, in press.

Darley, F., Efficacy of language rehabilitation in aphasia. Journal of Speech and Hearing Disorders, 37, 1972, 3-21.

Holland, A., Communicative Abilities in Daily Living (CADL): A Test of Functional Communication for Aphasic Adults. Baltimore: University Park Press, 1980.

Sarno, M.T. and Levita, E., Recovery in treated aphasia in the first year post-stroke. Stroke, 10, 6, 1979, 663-670.

Wertz, R., Language Rehabilitation in Aphasia: An Examination of the Process and Its Effects. Final Report, Veterans Administration Cooperative Study, 1978.

DISCUSSION

Q: Is CADL more sensitive to change or to different kinds of changes than are other tests?
A: I don't know.

Q: Does CADL ask people to do things during testing situation, whether in treatment or not, that other tests don't ask them to do?
A: Yes, I think so.

Q: In that sense, CADL then makes assumptions that other tests don't make?
A: Yes, that's the sense in which CADL gets at different things from other tests.

Q: Did you ask why they were discontinued? Because, so far as the design shows it, we don't know if the improvement is the result of treatment or if it's because the people who were not improving were discontinued.
A: I did not ask. There were five in the discontinued group who improved,

and one in the continued group who didn't, which I feel addresses the issue. There was no design, incidentally, these data all just "fell out."

Q: It still could be just an artifact of those discontinued people not showing much in the way of improvement.

A: I think I tried to say that traditional termination criteria probably were applied and in the long run, that the data say the discontinued patients were discontinued for fairly standard reasons. I think you can get that from the ranges; people who have made all the improvement they're going to make are there, as are the new global patients. It's not on the handout, but even the highest CADL scores were among the discontinued. It was also true on the BDAE scores where the whole range on its severity scale was represented.

Q: This can be viewed then as a validation of discontinuation criteria?

A: I'm suggesting that we do rather nicely by some of our discontinuation criteria, but I can't prove that with these data.

Q: Would you care to speculate on what happened in therapy that might have resulted in improvement in their abilities in daily living tasks?

A: I love to speculate--but in this case I would rather speculate about how nice it would have been to design this as a real study, and to have gathered data on the other tests as well.

Q: Regarding the patients who improved in the discontinued group, do you know anything about their activities, and about the people they interacted with, or anything else that might have contributed to their improvement?

A: I really don't. I know these patients fairly well, because we observed them intensively at the time of the first test in their natural environments. I tried to think back about this. To me, the most spectacular thing was the two oldest, institutionalized, most linguistically and socially impoverished patients, where I saw nothing at all during that observation to have predicted their improvement. There is nothing, in my recollection, to unify the causes of improvement for the whole group.

Q: I am interested in the possibility that some patients have conditions in their environment, observable from the very beginning, that aid treatment.

A: That's a terrifically important idea. I think that the environment that is naturally responsive and sensitive to the aphasic patient as well as the family who is trained to be responsive, aware of the patient's strategies and how to get him to use them more are tremendously important factors in facilitating recovery.

Q: Have you ever given the CADL twice within a two week period with no therapy?

A: Yes. That was how we originally checked its reliability, except that it was a three week period. The data are very convincing correlation of .99.

Q: There is a difference between statistical and clinical significance. would you say that this statistical improvement reflected clinical improvement?
A: The statistical improvement here is less compelling than many varieties of clinical improvement that are harder to document. I would love to say that I saw all kinds of clinical improvement, but I simply wasn't around these patients enough at the time of the second test to know that it occurred.

Q: Did the discontinued group include patients who never got treatment at all?
A: No.

Q: Did any folks in the discontinued group get <u>any</u> treatment between the first and second tests?
A: Yes. But I really don't know how many.

Q: Were any of them drop-outs? Did the clinician terminate them, and were they self-terminated?
A: I sure wish I could answer that. But I can't.

Q: I'm thinking the results could be more powerful because if therapy is good and some of the discontinued group got some of the good stuff, then that would imply that even though they got some of it, they didn't get as much as the other group and therefore they improved even more.
A: Could be.