



Pergamon

Archives of Clinical Neuropsychology, Vol. 10, No. 6, pp. 535-542, 1995
Copyright © 1995 National Academy of Neuropsychology
Printed in the USA. All rights reserved
0887-6177/95 \$9.50 + .00

0887-6177(95)00002-X

The Determination of Clinically Meaningful Cognitive Decline: Development and Use of an Alternative Method

David W. Desmond and Thomas K. Tatemichi

Department of Neurology, Columbia University

Yaakov Stern and Mary Sano

Departments of Neurology and Psychiatry and the Gertrude H. Sergievsky Center, Columbia University

Statistical methods traditionally used in the analysis of change (e.g., repeated measures ANOVA) may be inadequate for the investigation of cognitive decline if a study's effect size is small, the variance within groups is heterogeneous, or the statistical power is low. To examine an alternative approach to the determination of clinically meaningful cognitive decline and investigate whether such decline occurs during the first year after stroke, we administered a neuropsychological test battery to 172 patients (age = 70.3 ± 7.6 years; education = 10.3 ± 4.7 years) 3 and 12 months after stroke and 199 nondemented stroke-free control subjects (age = 71.1 ± 6.4 years; education = 12.8 ± 4.2 years) on two occasions 12 months apart. Two neuropsychologists classified each subject's test performance as having declined, improved, or remained stable based solely on clinical judgment. Reliability of the rating of decline versus the pooled rating of

Presented in part at the 22nd annual meeting of the International Neuropsychological Society, Cincinnati, OH, February, 1994. This work was supported in part by Grants R01-NS26179 and P01-AG07232 and Clinical Investigator Development Award K08-NS01635 (Dr. Desmond) from the National Institutes of Health.

Address correspondence and to Dr. Desmond, Neurological Institute, 710 West 168th Street, New York, NY 10032.

improvement/stability was excellent ($\kappa = 0.79$). The two rating groups differed significantly and in the appropriate directions in change on most tests. While a MANOVA comparing the stroke and control groups on change in test scores was not significant, logistic regression analysis determined that a rating of clinically meaningful cognitive decline was associated with stroke status (Odds Ratio = 1.8, 95% Confidence Interval = 1.0 to 3.2), while adjusting for demographic factors. We propose that this alternative approach to the analysis of cognitive change can facilitate the recognition of decline in subgroups of subjects. It would be valuable as an adjunct to studies of the incidence of dementia, for example, in which the recognition of cognitive decline might be difficult in highly educated patients whose baseline level of performance is far above the cutoffs operationalized for the diagnosis of dementia.

Statistical methods traditionally used in the analysis of change (e.g., repeated measures analysis of variance) are dependent on effect size, the homogeneity of variance within groups, and the statistical power of a study. This creates a potential problem for certain investigations because a statistically significant difference may not be clinically meaningful when large samples are examined while a clinically meaningful difference may not be statistically significant when small samples are examined. In addition, analytic methods that rely solely on comparisons based upon change in overall mean test scores may fail to recognize meaningful change within subgroups. We describe an alternative approach to the analysis of change based upon neuropsychologists' judgments of clinically meaningful cognitive decline and present data regarding cognitive function in the first year after stroke to illustrate the utility of this approach.

METHOD

Subjects

As part of a prospective study of stroke and dementia (Tatemichi et al., 1992), we examined 172 patients (age = 70.3 ± 7.6 years; education = 10.3 ± 4.7 years) 3 and 12 months after stroke and 199 nondemented stroke-free control subjects (age = 71.1 ± 6.4 years; education = 12.8 ± 4.2 years) on two occasions 12 months apart. Stroke patients were ineligible for this study if they had received a diagnosis of possible Alzheimer's disease with stroke based upon baseline test results consistent with dementia and a prestroke history of functional impairment. The control sample was composed of subjects who were randomly selected from Medicare records and a commercial list, volunteers from the community, and spouses of stroke patients. Note that we based our analyses on subjects with and without stroke to validate a new method for the determination of clinically meaningful cognitive decline and not to investigate the correlates of cognitive decline after stroke.

Test Instruments

We administered a battery of neuropsychological tests (Stern et al., 1992) to all subjects. This battery included measures of verbal and nonverbal memory (i.e., the Selective Reminding Test and a multiple-choice recognition version of the Benton Visual Retention Test), orientation (i.e., the Mini-Mental State Examination orientation subtest), verbal functions (i.e., a 15-item version of the Boston Naming Test, letter and category fluency tests, and the repetition and complex ideation subtests of the Boston Diagnostic Aphasia Examination), visuospatial functions (i.e., the Rosen Drawing Test and a multiple-choice matching version of the Benton Visual Retention Test), and verbal and nonverbal abstract reasoning skills (i.e., the Similarities subtest of the Wechsler Adult Intelligence Scale-Revised and the Identities and Oddities subtest of the Mattis Dementia Rating Scale). This battery was developed for use in epidemiologic studies of dementia; thus, we recognize that it may not be optimal for the comprehensive assessment of cognitive function in patients following stroke.

Procedures

The neuropsychological test battery was administered to all subjects at both examination points and an alternate word list was used for the Selective Reminding Test during the second examination to minimize any potential learning effect. Two experienced clinical neuropsychologists (Y.S., M.S.) reviewed the neuropsychological test scores received by each subject during the two examinations, discussed each case, and, using only clinical judgment, came to consensus on whether each subject's test scores had improved, declined, or remained stable. Although the raters were not blind to clinical information, they were asked to disregard such information in determining their clinical change ratings.

Statistical Analyses

In all analyses, subjects who were considered to have declined were compared to the pooled group of those who had improved and remained stable. To determine whether the change ratings were reliable, the test results of 30 stroke patients and 30 control subjects were randomly selected for rereview and kappa coefficients were generated by comparing the initial ratings with the reratings. We then performed a stepwise discriminant analysis, with change rating group as the dependent variable and the differences between neuropsychological subtest scores received at the first and second examinations as independent variables, to determine the grouping of change scores that best discriminated between the change rating groups (i.e., the group of tests that was most important to the raters in formulating the clinical change ratings). Only those difference scores that were found to be related to change

rating group in *t*-tests ($p < .10$) were entered into the discriminant analysis. We also compared the two groups on change in scores on cognitive and functional measures that were independent of the change ratings, including the Mini-Mental State Examination; the Blessed Functional Activity Scale, which assesses the cognitive aspects of functional disability; the Barthel Index, which assesses the physical aspects of functional disability; and the Stroke Severity Scale, which assesses the severity of neurologic impairment. These measures are described in more detail in another publication (Tatemichi et al., 1992). The recognition of greater decline on these independent measures in subjects who are rated as having declined on the neuropsychology battery would provide an additional level of validation for the change rating procedure.

To illustrate a potential use of these change ratings, we investigated stroke status as a correlate of cognitive decline. We first utilized a traditional approach to the analysis of change, a repeated measures MANOVA, in comparing the stroke group to the control group with regard to their neuropsychological subtest scores at each of the two examinations. We then performed a logistic regression analysis, with change rating as the dichotomous dependent grouping variable and stroke status and demographic factors (i.e., age, trichotomized as 80+, 70–79, or 60–69 years; education, dichotomized as < 12 or 12+ years; race/ethnicity, coded as white, black, or Hispanic; and sex) as independent variables to investigate the role of stroke as a correlate of clinically meaningful cognitive decline.

RESULTS

Reliability of the Change Ratings

The reliability analysis resulted in a kappa coefficient of 0.79. Fifty-six of the 60 reratings (93.3%) were consistent with the original ratings.

Test Performance of the Change Rating Groups

Univariate analyses determined that the two rating groups differed significantly in neuropsychological subtest score change, with the direction of test score change consistent with the clinical change rating, on most of the subtests. These results are presented in Table 1. Discriminant analysis determined that change on the following neuropsychological measures best discriminated between the change rating groups: Selective Reminding Test total recall and delayed recall and recognition, the Benton Visual Retention Test recognition task, orientation, category fluency, the Rosen Drawing Test, Similarities subtest raw score, and the Identities and Oddities subtest. Use of this grouping of tests in a discriminant function classified subjects into the two change rating groups with a sensitivity of 0.94 and a specificity of 0.85.

When the two rating groups were compared with regard to change in scores on cognitive and functional measures that were independent of the change rat-

TABLE 1
Neuropsychological Test Score Change by Rating Group

Test	Improved/Stable (<i>n</i> = 300)	Decline (<i>n</i> = 71)	<i>p</i>
Selective Reminding Test			
Total Recall	2.9 ± 6.5	-6.2 ± 5.9	< 0.0001
Long-Term Recall	5.1 ± 9.6	-6.6 ± 8.7	< 0.0001
Delayed Recall	0.3 ± 1.9	-1.8 ± 1.9	< 0.0001
Delayed Recognition	-0.2 ± 1.6	-1.2 ± 1.9	< 0.0001
BVRT — Recognition	0.2 ± 1.7	-0.2 ± 1.8	0.0687
Orientation	0.1 ± 1.0	-0.6 ± 1.8	< 0.0001
Boston Naming Test (15-item)	0.3 ± 1.3	-0.2 ± 1.9	0.0098
Verbal Fluency — CFL (mean)	0.4 ± 2.3	-0.8 ± 2.4	0.0001
Categories (mean)	0.2 ± 2.9	-1.5 ± 3.4	< 0.0001
Repetition	0.0 ± 0.7	0.0 ± 0.9	0.8512
Complex Ideation	0.0 ± 0.9	-0.3 ± 1.0	0.0090
Rosen Drawing Test	-0.1 ± 0.9	-0.6 ± 1.1	0.0001
BVRT — Matching	0.2 ± 1.4	0.2 ± 2.1	0.9424
Similarities — Raw Score	-0.2 ± 3.5	-1.9 ± 3.9	0.0004
Identities and Oddities	0.2 ± 1.7	-0.6 ± 2.4	0.0014

Note. Figures presented are means ± 1 standard deviation and represent the score at the 12-month examination minus the score at the 3-month examination. Significance levels are based on *t*-tests. BVRT = Benton Visual Retention Test.

ings, it was found that they differed significantly and in the appropriate directions in change on the Mini-Mental State Examination and the Blessed Functional Activity Scale but not on the Barthel Index or the Stroke Severity Scale. These results are presented in Table 2.

Stroke as a Correlate of Cognitive Decline

The repeated measures MANOVA comparing the stroke and control groups on change on the battery of neuropsychological tests was not significant [$F(15, 330) = 0.95, p = .512$]. Logistic regression analysis, however, determined that clinically meaningful cognitive decline was significantly related to stroke status [Odds Ratio (OR) = 1.8, 95% Confidence Interval (CI) = 1.0 to 3.2] and age 80+ versus 60–69 (OR = 2.1, CI = 1.2 to 3.8), but not age 70–79 versus 60–69, race, or sex. Thirty-nine of the 172 stroke patients (22.7%) were rated as having declined, while 32 of the 199 control subjects (16.1%) received that rating.

DISCUSSION

Our results suggest that neuropsychologists' judgments of clinically meaningful cognitive decline are reliable (Fleiss, 1981) and consistent with change that can be observed in neuropsychological test scores. Discriminant analysis determined that the neuropsychologists relied upon a variety of subtests tapping all

TABLE 2.
Change in Other Test Scores by Rating Group

Test	Improved/Stable (<i>n</i> = 300)	Declined (<i>n</i> = 71)	<i>p</i>
Mini-Mental State Examination	0.1 ± 2.2	-1.4 ± 3.6	< 0.0001
Blessed Functional Activity Scale	0.3 ± 1.2	0.8 ± 2.6	0.0279
Barthel Index	0.1 ± 9.1	-0.9 ± 17.6	0.5249
Stroke Severity Scale	0.2 ± 2.1	0.2 ± 2.4	0.9494

Note. Figures presented are means ± 1 standard deviation and represent the score at the 12-month examination minus the score at the 3-month examination. Significance levels are based on *t*-tests.

cognitive domains in formulating the change ratings. The rating groups differed significantly in change on the Mini-Mental State Examination and the Blessed Functional Activity Scale, which tap cognitive function and the cognitive aspects of activities of daily living, respectively, but not on the Barthel Index or Stroke Severity Scale, which tap the physical aspects of neurologic impairment, suggesting that the change ratings were specific to cognitive function.

Although some investigators have suggested that an actuarial approach (i.e., the use of an operationalized diagnostic paradigm) is more reliable and valid than clinical judgment (Dawes, Faust, & Meehl, 1989), others have emphasized the utility of clinical judgment in neuropsychological practice (Butters et al., 1990). Some previous studies have shown that the clinical judgment of neuropsychologists is both reliable and sensitive to physiologic measures of disease severity (Grant, Heaton, McSweeney, Adams, & Timms, 1982; Heaton, Grant, Anthony, & Lehman, 1981; Heaton, Grant, McSweeney, Adams, & Petty, 1983), and it has served as the basis for a computerized algorithm designed to facilitate the recognition of worsening in neuropsychological test performance in patients with diabetes (DCCT Research Group, 1994).

While the use of a traditional analytic technique (i.e., repeated measures MANOVA) did not permit the recognition of a significant difference between the stroke and control groups in test score change, our alternative approach found that patients may exhibit clinically meaningful cognitive decline during the first year after stroke. Numerous cross-sectional studies have demonstrated that stroke frequently results in cognitive impairment (Hom & Reitan, 1990; Tatemichi et al., 1994). Most previous prospective studies of cognitive function after stroke have suggested that patients who survive tend to exhibit improvement (Wade, Skilbeck, & Hewer, 1989) and that recovery of function usually plateaus after three months (Skilbeck, Wade, Hewer, & Wood, 1983). Those studies that have suggested that decline may occur following stroke have focused on the incidence of dementia, the risk of which has been shown to be significantly increased in patients with stroke (Aronovich, Treves, Bornstein, & Korczyn, 1992; Kotila, Waltimo, Niemi, & Laaksonene, 1986; Tatemichi et al., 1993).

Ratings of clinically meaningful cognitive change can be used to identify individual subjects who have declined and who can then be pooled into subgroups that might not be recognized if more traditional analytic techniques are used. Future studies should utilize a similar technique if decline is suspected in small subgroups or efforts are being made to identify subtle risk factors. This alternative approach might be particularly valuable as an adjunct to studies of the incidence of dementia, in which the recognition of cognitive decline might be difficult in highly educated patients whose baseline level of performance is far above the cutoffs operationalized for the diagnosis of dementia.

REFERENCES

- Aronovich, B. D., Treves, T. A., Bornstein, N. M., & Korczyn, A. (1992). Dementia after first stroke: 3-years survival analysis [abstract]. *Cerebrovascular Diseases*, *2*, 216.
- Butters, N., Grant, I., Haxby, J., Judd, L. L., Martin, A., McClelland, J., Pequegnat, W., Schacter, D., & Stover, E. (1990). Assessment of AIDS-related cognitive changes: Recommendations of the NIMH Workshop on Neuropsychological Assessment Approaches. *Journal of Clinical and Experimental Neuropsychology*, *12*, 963–978.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668–1674.
- DCCT Research Group. (1994). A screening algorithm to identify clinically significant changes in neuropsychological functions in the Diabetes Control and Complications Trial. *Journal of Clinical and Experimental Neuropsychology*, *16*, 303–316.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.
- Grant, I., Heaton, R. K., McSweeney, A. J., Adams, K. M., & Timms, R. M. (1982). Neuropsychologic findings in hypoxicemic chronic obstructive pulmonary disease. *Archives of Internal Medicine*, *142*, 1470–1476.
- Heaton, R. K., Grant, I., Anthony, W. Z., & Lehman, R. A. W. (1981). A comparison of clinical and automated interpretation of the Halstead-Reitan battery. *Journal of Clinical Neuropsychology*, *3*, 121–141.
- Heaton, R. K., Grant, I., McSweeney, A. J., Adams, K. M., & Petty, T. L. (1983). Psychologic effects of continuous and nocturnal oxygen therapy in hypoxicemic chronic obstructive pulmonary disease. *Archives of Internal Medicine*, *143*, 1941–1947.
- Hom, J., & Reitan, R. M. (1990). Generalized cognitive function after stroke. *Journal of Clinical and Experimental Neuropsychology*, *12*, 644–654.
- Kotila, M., Waltimo, O., Niemi, M.-L., & Laaksonene, R. (1986). Dementia after stroke. *European Neurology*, *25*, 134–140.
- Skilbeck, C. E., Wade, D. T., Hewer, R. L., & Wood, V. A. (1983). Recovery after stroke. *Journal of Neurology, Neurosurgery, and Psychiatry*, *46*, 5–8.
- Stern, Y., Andrews, H., Pittman, J., Sano, M., Tatemichi, T., Lantigua, R., & Mayeux, R. (1992). Diagnosis of dementia in a heterogeneous population: Development of a neuropsychological paradigm-based diagnosis of dementia and quantified correction for the effects of education. *Archives of Neurology*, *49*, 453–460.
- Tatemichi, T. K., Desmond, D. W., Mayeux, R., Paik, M., Stern, Y., Sano, M., Remien, R., Williams, J. B. W., Mohr, J. P., Hauser, W. A., & Figueroa, M. (1992). Dementia after stroke: Baseline frequency, risks, and clinical features in a hospitalized cohort. *Neurology*, *42*, 1185–1193.
- Tatemichi, T. K., Desmond, D. W., Stern, Y., Paik, M., Sano, M., & Bagiella, E. (1994). Cognitive impairment after stroke: Frequency, patterns, and relationship to functional abilities. *Journal of Neurology, Neurosurgery, and Psychiatry*, *57*, 202–207.

- Tatemichi, T. K., Paik, M., Bagiella, E., Desmond, D. W., Figueroa, M., Gropp, T. I., Stern, Y., Sano, M., Hauser, W. A., & Mayeux, R. (1993). Risk of dementia after stroke: Incidence and antecedents [abstract]. *Neurology*, **43** (Supplement 2), A277.
- Wade, D. T., Skilbeck, C. E., & Hower, R. L. (1989). Selected cognitive losses after stroke: Frequency, recovery and prognostic importance. *International Disability Studies*, **11**, 34–39.