

Test-Retest Stability of the Repeatable Battery for the Assessment of Neuropsychological Status in Schizophrenia

Christopher M. Wilk, M.A.

James M. Gold, Ph.D.

John J. Bartko, Ph.D.

Faith Dickerson, Ph.D.

Wayne S. Fenton, M.D.

Michael Knable, D.O.

Christopher Randolph, Ph.D.

Robert W. Buchanan, M.D.

Objective: The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) was designed as a cognitive screening test, providing both a total scale score and five specific cognitive ability index scores. This study examined the test-retest stability of the RBANS in individual patients with schizophrenia relative to a healthy comparison group.

Method: A total of 181 patients with schizophrenia or schizoaffective disorder were recruited from three clinical settings. Healthy comparison subjects were recruited as part of the RBANS standardization. Participants were administered one form of the RBANS on one occasion and another form at a later date, with intervals ranging from 1 to 134 days.

Results: Intraclass correlation coefficients for the RBANS total scale were 0.84

for the patients with schizophrenia and 0.77 for the healthy comparison subjects. Confidence intervals and percentile data for the total scale change scores were similar for both groups.

Conclusions: The RBANS demonstrated reasonable intraclass correlation coefficient test-retest reliability for both schizophrenia patients and healthy comparison subjects. Confidence intervals are comparable to those previously published for the WAIS-R and Wechsler Memory Scale—Revised, suggesting that retest measurement error is not dramatically increased in the RBANS, despite the brevity of the test. These data may serve as an informative guide for using the RBANS to evaluate neuropsychological change on the level of the individual subject.

(*Am J Psychiatry* 2002; 159:838–844)

Cognitive impairment is a hallmark feature of schizophrenia and is evident across many cognitive domains. Deficits of learning, memory, and attention are the most reliable neuropsychological findings (1–3). Impairments of executive function, visuospatial abilities, and language have also been associated with the illness (4, 5). These cognitive deficits are present at the onset of the illness, tend not to be related to variations in the illness over time, are minimally related to symptom severity (6, 7), and tend not to be ameliorated by traditional pharmacological therapies (8), but are related to functional outcome (9). As such, cognitive impairment in patients with schizophrenia is pervasive across a number of domains, and assessment of these domains is often indicated in clinical settings.

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) is a standardized screening instrument designed to assess global neuropsychological functioning in a brief administration. This instrument measures several cognitive domains of interest in schizophrenia—immediate memory, visuospatial/constructional ability, language, attention, and delayed memory—and provides a global measure, the total scale score (10). In addition, the RBANS offers two alternate forms to reduce the potential influence of practice effects in serial test administration.

In the initial reports on the use of the RBANS with patients with schizophrenia, Gold and colleagues (11) demonstrated that the test was sensitive to impairments typically found in schizophrenia. In a group of 129 patients, the mean RBANS total score was 71.4, nearly two full standard deviations ($SD=15.0$) below the normal mean reported in the RBANS manual. Further, RBANS performance was strongly related to functional outcome: competitively employed patients had a mean RBANS total scale score of 86.8 ($SD=12.6$), whereas unemployed patients had a mean score of 70.8 ($SD=19.3$). In the total study group, language and visuospatial functions were relatively spared, compared with more severe impairments of memory and attention. RBANS scores were relatively independent of symptom severity and were highly correlated with scores on the WAIS-III and Wechsler Memory Scale, 3rd ed. (WMS-III), but were less associated with measures of executive function, motor performance, and vigilance (12). These data clearly indicate that the RBANS does not provide a comprehensive neuropsychological profile (see references 11 and 12 for discussion of this issue), but the brevity of the test may enhance its utility in a variety of clinical settings. In a smaller study group over a limited time interval, the RBANS exhibited good alternate-form test-retest reliability, with promising actual score and relative posi-

tion stability, and lacked any indication of a practice effect (11). Thus, the preliminary report based on analyses of group data suggested that the RBANS might be a useful tool for evaluating changes in cognitive functioning related to changes in treatment or clinical state in individuals with schizophrenia.

There are several issues that need to be considered in evaluating the adequacy of the RBANS as a measure of change observed in individual subjects. As noted by McCaffrey and colleagues (13), only the reliability coefficients are reported for the vast majority of psychometric instruments. Also, inappropriate measures of association, such as Pearson's correlation, are sometimes used to measure reliability. However, such measures of association do not necessarily assess agreement between ratings. For example, time-2 measures that are uniformly larger than time-1 measures will produce a large Pearson correlation but obviously cannot and do not reflect observed-score agreement. In contrast, the intraclass correlation coefficient (ICC) (14) assesses the degree of variation from time-1 to time-2 values. If these differences are small, ICC reliability will be high; large differences will yield a low ICC. Thus, in assessing true agreement, one should use the ICC as opposed to simple measures of association (e.g., Pearson's *r*).

In addition, it is often assumed that measures for individual patients behave in the same manner as the mean of that patient's clinical group. For example, although groups tested twice may obtain similar scores on the two test occasions, test-retest scores for the individuals in the group may fluctuate substantially (15, 16). Matarazzo and Herman (17) examined the test-retest stability of the WAIS-R in 119 nonclinical subjects. They found that, despite the extremely high test-retest reliability coefficients for full-scale IQ reported in the manual (0.96 and 0.97) (18) and a relatively small difference in group test-retest score (6.2 points), individual subjects exhibited test-retest change scores ranging from losses of 12 points to gains of 20 points. Such a wide range was not simply an artifact of anomalous extreme values: the standard deviation of the change scores was 5.07 scaled score points for full-scale IQ, which yields 90% confidence limits of -2.18 (lower) and 14.63 (upper). However, assuming the same sample size and standard deviation, only a 0.07-point decline or a 3.45-point gain would be required for statistical difference at the group level. Thus, when interpreting observed differences in scores, it is important to consider base-rate data rather than to extrapolate from group mean change scores.

The purpose of the study reported here was to examine the test-retest stability of the two RBANS forms over an extended period of time in a large group of patients with schizophrenia or schizoaffective disorder relative to a healthy comparison group. Specifically, our goals were to provide 1) ICC reliability coefficients for each of the RBANS cognitive index scores, 2) base-rate data for change in test-retest total scale scores, and 3) total scale confidence inter-

vals based on estimates of variability due to random measurement error. The base-rate data illustrate the percentage of patients with schizophrenia obtaining a given RBANS total scale change score. Confidence intervals are useful for assessing the degree of certainty that a given difference in total scale scores is not due to random measurement error. Such data may serve as an informative guide for evaluating individual neuropsychological change in patients with schizophrenia or schizoaffective disorder. This descriptive approach is appropriate for a test with published alternate forms and takes advantage of the large size of the study group in the analysis. Other recent approaches discussed in the neuropsychological literature are designed for tests lacking alternate forms or where the goal is highly precise predictions of retest score on the basis of complex regression models (15, 19–25, unpublished 1998 study of C.J. Chelune). The statistical approach and data presentation that follow were designed to inform everyday clinical decision-making.

Method

Subjects

Two patient groups totaling 181 participants were included in the analysis. All patients met the DSM-IV criteria for schizophrenia or schizoaffective disorder and did not meet the criteria for alcohol/drug dependence or mental retardation. Data for group 1 (75 outpatients, 24 inpatients) were collected at the Maryland Psychiatric Research Center. These patients (71 men, 28 women) were between the ages of 18 and 60 years (mean=39.89, SD=9.36) and had a mean of 11.71 years (SD=2.60) of education; 33 patients had less than 12 years of education, 36 patients had 12 years of education, 22 patients had attended but did not complete college, and eight patients completed 16 or more years of education. Data for group 2 (82 outpatients) were collected at the Sheppard Pratt Health System and Chestnut Lodge Hospital. These patients (51 men, 31 women) were between the ages of 20 and 64 years (mean=40.71, SD=9.70). Of these patients, seven had less than 12 years of education, 23 had 12 years of education, 24 had attended but did not complete college, and 27 patients completed 16 or more years of education (educational data were missing for one subject).

Data for 99 healthy comparison subjects were obtained from the RBANS standardization sample (10). These participants (28 men, 71 women) were between the ages of 24 and 86 years (mean=64.46, SD=14.76) and had a mean of 15.43 years (SD=2.83) of education. No comparison participant had a history of stroke, seizure, or central nervous system infection or disease, or met the criteria for a major psychiatric illness.

Materials and Procedure

All participants received the alternate forms of the RBANS on two separate occasions. Patients in group 1 were approached after they were judged by their clinicians to be clinically stable. After providing written informed consent, 88 patients received Form A followed by Form B and 11 received the reverse order. The number of days between testing ranged from 14 to 134, with a mean of 50.65 (SD=27.28). Of these 99 participants, 59 remained on the same medication at the same dose during and between testing occasions and 40 experienced medication adjustment (type and/or dose change). Using a *t* test for independent groups, we observed no difference in total scale test-retest change scores

TABLE 1. Scores on the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) of Schizophrenic Patients and Healthy Comparison Subjects on Two Test Occasions and Test-Retest Change in Scores^a

Group and RBANS Measure	Score on Test Occasion 1		Score on Test Occasion 2		Change	
	Mean	SD	Mean	SD	Mean	SD
Schizophrenic patients (N=181)						
Total scale	71.81	14.92	73.50	8.56	1.69 ^b	8.56
Cognitive index						
Immediate memory	72.18	18.04	76.08	14.62	3.90 ^b	14.62
Visuospatial/constructional ability	80.57	17.92	79.34	13.66	-1.23	13.66
Language	84.73	14.85	88.02	13.41	3.29 ^b	13.41
Attention	75.43	17.44	76.50	11.06	1.07	11.06
Delayed memory	74.00	18.72	74.91	15.22	0.91 ^b	15.22
Healthy comparison subjects (N=99)						
Total scale	106.30	13.84	104.82	8.97	-1.48 ^b	8.97
Cognitive index						
Immediate memory	105.75	12.64	105.77	12.29	0.02 ^b	12.29
Visuospatial/constructional ability	99.42	13.22	100.42	12.63	1.00	12.63
Language	105.69	14.23	101.85	16.25	-3.84 ^b	16.25
Attention	106.32	14.33	108.52	9.34	2.20	9.34
Delayed memory	105.38	13.93	100.95	13.86	-4.43 ^b	13.86

^a Range of test-retest interval=1–134 days.^b Significant difference between groups (t tests for independent groups, df=278, p<0.05).

between patients who had a medication adjustment and those who did not ($t=-1.03$, $df=97$, $p=0.30$). The data from these two groups were combined for further analyses.

Data for group 2 were collected under the auspices of a 16-week double-blind clinical trial of omega-3 fatty acid (eicosapentaenoic acid) supplements conducted at the Sheppard Pratt Health System and Chestnut Lodge Hospital. These patients were assessed to be clinically stable, and they provided written informed consent before entry into the study. Adjunctive to receiving their current antipsychotic medication regimen, patients were randomly assigned to receive either eicosapentaenoic acid or a mineral oil placebo. Both the intervention and the placebo were found to have inert effects on cognition. The eicosapentaenoic acid group had a mean total scale score of 75.12 at baseline with 76.07 at follow-up (a mean change of 0.95), whereas the placebo group had a mean total scale score of 70.68 at baseline compared with 73.71 at follow-up (a mean change of 3.03) (25). All participants in this group received Form A followed by Form B.

T tests for independent groups were performed on the Form A-Form B change scores to establish comparability of the Maryland Psychiatric Research Center and Sheppard Pratt Health System/Chestnut Lodge Hospital patient groups. No statistically significant differences were revealed between the groups for any of the RBANS index or total scale change scores. Furthermore, the distributions of the change scores closely approximated a normal distribution for both groups (all data passed omnibus as well as skewness and kurtosis tests for normality). Consequently, the two groups were combined for a total of 181 patients.

The two RBANS forms were administered in a counterbalanced design for the standardization comparison participants: 52 participants received Form A followed by Form B, and 47 received Form B followed by Form A. The test-retest interval ranged from 1 to 7 days (10). We compared the test-retest total scale scores of these two healthy comparison subgroups and unexpectedly found evidence of an order effect for Form B. Specifically, subjects tested on Form A first had a mean total scale score of 104.50 ($SD=12.43$) followed by a Form B score of 105.19 ($SD=12.22$), yielding a mean Form A-Form B total scale score change of 0.69 points ($SD=7.94$). Subjects tested with Form B first had a mean total scale score of 108.28 ($SD=15.15$) followed by a Form A score of 104.40 ($SD=13.98$), yielding a total scale score change of -3.87 points ($SD=9.51$). A repeated measures analysis of variance revealed a statistically significant interaction between test form and test oc-

casion, with subjects receiving Form B first having a lower score on retest with Form A ($F=6.76$, $df=1, 98$, $p=0.01$). A Bonferroni post hoc t test using the mean square error for testing across time within the group receiving Form B first revealed a significant difference ($t=3.05$, $df=194$, $p<0.01$). Thus, the major component of this interaction was explained by the change from test occasion 1 to test occasion 2 within the group receiving Form B first (the analysis of the effect of Form A versus Form B on the first testing occasion was not statistically significant). We have no plausible explanation for this pattern of results. We considered possible nonspecific practice effects, subtle differences in difficulty across test forms, and the potential interaction of these two factors, but we could not produce a logical explanation for why Form B scores should be slightly higher only when this form was administered first. Thus, this appears to be a chance finding, and we have chosen to combine the two test order groups in the analyses reported below. Furthermore, although this test-order effect was statistically significant, the width of the test-retest confidence intervals given in the Results section illustrates the clinical "nonsignificance" of this result.

Data Analysis

Change scores for the total scale and each cognitive index were obtained by subtracting the observed score at time 1 from the observed score at time 2. T tests for independent groups were used to compare the groups on each of these change scores. Pearson's r was used as an index to assess changes in the relative rank of scores between time 1 and time 2, and the ICC was used as an index of observed-score agreement between these test scores. Statistical confidence limits were obtained for the total scale change score.

Results

The RBANS demonstrated clear sensitivity to the type of impairment observed in patients with schizophrenia: this study group of 181 patients had a mean total score of 71.81 ($SD=14.92$), nearly two standard deviations below the normal mean, very similar to the level of performance observed in our initial report (11). The RBANS pretest, posttest, and change scores for the patients and the healthy comparison subjects are presented in Table 1. The actual

TABLE 2. Association Between and Reliability of Test-Retest Scores on the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) of Schizophrenic Patients and Healthy Comparison Subjects

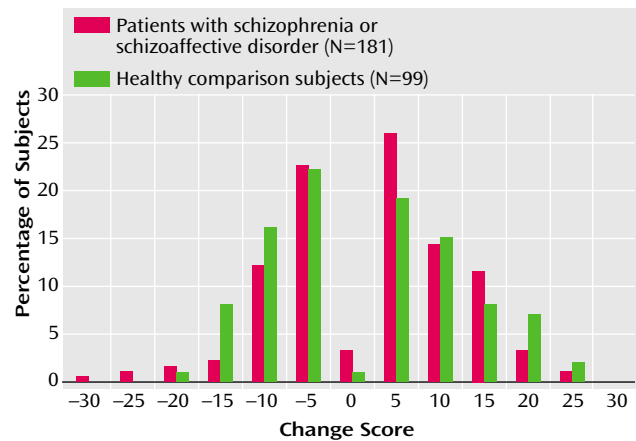
RBANS Measure	Measure of Stability of Scores From Test to Retest ^a			
	Schizophrenic Patients (N=181)		Comparison Group (N=99)	
	Association (r)	Reliability (ICC)	Association (r)	Reliability (ICC)
Total scale	0.84	0.84	0.78	0.77
Cognitive index				
Immediate memory	0.69	0.67	0.55	0.55
Visuospatial/constructual ability	0.71	0.71	0.53	0.54
Language	0.54	0.51	0.38	0.36
Attention	0.81	0.81	0.77	0.76
Delayed memory	0.69	0.68	0.57	0.53

^a ICC=intraclass correlation coefficient.

mean differences in change scores were quite small (none were greater than 4 points) and thus were not likely to be clinically significant in light of the magnitude of the differences in the comparisons of the patients and the healthy subjects (approximately 30 points). Some of the test-retest change scores were significantly different between groups (e.g., total scale [$t=2.91$, $df=278$, $p=0.004$], immediate memory [$t=2.24$, $df=278$, $p<0.03$], language [$t=3.94$, $df=278$, $p=0.001$], and delayed memory [$t=2.9$, $df=278$, $p=0.004$]). However, the standard deviations of the change scores were strikingly similar between the two groups. For example, on the total scale, the standard deviation was 8.56 in patients and 8.97 in the comparison group. This suggests that the RBANS behaves in a remarkably similar manner in both groups across repeated testing occasions.

Table 2 presents the measures of association (Pearson's r) and agreement-reliability (ICC) for each of the index scores for both groups. Despite having a markedly longer interval between test occasions, the patients tended to have a slightly higher measure of alternate-form test stability than the healthy comparison group. Within the groups, the Pearson r and ICC values were quite similar for each score, indicating that the members of each group tended to retain their absolute score as well as their relative rank across testing occasions. Thus, practice effects and "test sophistication" did not appear to be operative. In fact, some mean index scores were lower on the second test occasion (Table 1). Overall, the total scale demonstrated the highest capacity for stable measurement across testing occasions.

Since the total scale yielded the highest test-retest stability and is a more global measure of cognitive functioning, we chose the total scale score as a basis for presenting base-rate data and calculating confidence intervals. Figure 1 presents the percentage of subjects in each group obtaining a given total scale change score. It is noteworthy that the distributions of the change scores were exceptionally

FIGURE 1. Distribution of Total Scale Test-Retest Change in Scores on the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) of Patients With Schizophrenia or Schizoaffective Disorder and Healthy Comparison Subjects^a

^a A change score of exactly 0 represents no test-retest difference. Each bar includes the change score with which it is labeled and the scores that fall between the previous and current values (e.g., 5 represents values >0 up to and including 5; 10 represents values >5 up to and including 10).

similar for both groups. Although a very small percentage of subjects exhibited no change at all, more than 50% of the patients had differences in scores of ± 5 points. Among the healthy comparison subjects, approximately 42% had differences of ± 5 points. Seventy-eight percent of the patients were within ± 10 points, and 73% of the healthy comparison subjects fell within this range.

The tails of this distribution provide the frequency of a given increase or decrease in total scale change scores. For example, by adding the values at 10 and below (through -30), one can see that 83.96% of the patients with schizophrenia obtained a change score of 10 points or less (82.82% of the healthy comparison subjects showed the same rate of change). Conversely, if one wanted to determine the percentage of subjects whose scores changed by -5 points or greater, one could add the percentages from -5 through 30 (in this case, 82.3% of patients showed this rate of change versus 74.74% of healthy comparison subjects). Such base-rate data provides a tool for the clinician to assess the rarity of a given change score.

Confidence limits for the total scale change scores for patients with schizophrenia and the healthy comparison group are presented in Table 3. Overall, the confidence intervals tended to be large, requiring more than a 10-point change in either direction to establish even a 90% confidence interval. However, this was true for healthy comparison subjects as well as for patients. Moreover, any given confidence interval was slightly larger for the healthy comparison subjects than for the patients with schizophrenia. This result is likely a function of the smaller size of the comparison group.

TABLE 3. Confidence Intervals for Test-Retest Change in Total Scores on the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) of Schizophrenic Patients and Healthy Comparison Subjects

Level of Confidence	Confidence Interval for Schizophrenic Patients (N=181)		Confidence Interval for the Comparison Group (N=99)	
	Lower Limit	Upper Limit	Lower Limit	Upper Limit
99%	-20.6	24.0	-25.1	22.1
95%	-15.2	18.6	-19.3	16.3
90%	-12.5	15.8	-16.4	13.4
85%	-10.7	14.1	-14.5	10.4
80%	-9.3	12.7	-13.1	10.1
70%	-7.2	10.6	-7.9	10.8

Discussion

The results of this study indicate that the RBANS is a potentially useful instrument for evaluating neuropsychological change in patients with schizophrenia. Alternate-form stability coefficients for patients with schizophrenia obtained over longer test-retest intervals (i.e., a mean of 7 weeks in group 1; 16 weeks in group 2) were remarkably comparable to (in fact slightly higher than) reliability coefficients for the healthy comparison subjects obtained over an interval of 1 week or less. Also, the ICCs were comparable to the Pearson coefficients, indicating that the stability of the relative rank of scores across testing occasions was driven by absolute score stability.

The RBANS index scores, however, were not equally stable. The language index stability coefficient was relatively low for the patients with schizophrenia and even lower for the healthy comparison subjects. This is likely a consequence of the language index score variability being driven primarily by only one subtest (speeded verbal fluency). However, the total scale score was quite promising, with a stability coefficient rivaling that reported for the WMS-III general memory index (for the RBANS, stability coefficient=0.84 for patients and 0.77 for comparison subjects; for the WMS-III, stability coefficient=0.87) (26). The attention index was also highly stable (stability coefficient=0.81 for patients and 0.76 for comparison subjects), approaching that of the WAIS-III processing speed index (stability coefficient=0.88) (26). The RBANS stability coefficients were impressive, in light of the fact that the WAIS-III and the WMS-III can each take three times as long as the RBANS to administer. On the basis of these data, the RBANS is a viable candidate for evaluating neuropsychological change in longer-term interventions.

When evaluating the differences between an individual's scores obtained on two separate occasions, practitioners can approach clinical decisions in two different ways. The first is by examining the base rate for change scores in a similar diagnostic group. This approach specifies how rare (or common) it is for a given change in scores to occur in the reference group. For example, a practitioner might wish to know how common it is for a patient with schizophrenia to have a decrease of more than 10 points on the

RBANS total scale when the patient is retested in the absence of any intervention (the inverse approach can be applied to improvement). If the percentages of subjects with more than a 10-point decline (Figure 1) are added, it can be noted that such a decrease occurred in fewer than 5.52% of the patients with schizophrenia in the study (and fewer than 9.09% of the healthy comparison subjects). In determining the level of improvement in patients, an increase of greater than 10 points was also rare (16.01% of subjects exhibited this degree of change in our study group). This level of increase is nearly two-thirds the 16-point difference between the mean total scale scores for unemployed and employed patients reported by Gold et al. (11). Indeed, retest improvements of 15 points occurred in only 4.41% of patients, suggesting that this degree of change is unlikely to occur by chance and is likely to be clinically significant. In fact, an improvement of 15.8 points defined the upper limit of the 90% confidence interval, an uncanny identity between observed change score distributions, confidence intervals, and clinical criterion validity data.

Another approach to evaluating neuropsychological change is to establish limits for a "statistically significant" difference in scores versus a difference based on measurement error. Confidence intervals can be used to accomplish this end. For the patients with schizophrenia and for the healthy comparison group in this study, the confidence intervals tended to be large: at 90% confidence, the distance between the lower limit and the upper limit for the total scale score was 28.3 for patients and 29.8 for comparison subjects, indicating that a greater change in a test score is required if the practitioner wishes to be assured that such change is not the result of random measurement error. The differences between the limits for the healthy comparison subjects and the patients with schizophrenia for any given confidence interval were trivial (an average difference of 1.4 points), further suggesting that the stability of the RBANS is similar in these two groups. The dramatic difference between decision making on the basis of group versus individual data is evident: with group data, an observed mean change of 1.69 points was significant at the 0.05 level of confidence; however achieving this level of confidence at the individual level would require a substantially larger change in score.

Although the RBANS confidence intervals may appear large, other well-respected, commonly used neuropsychological measures have comparable intervals. The WAIS-R, which is the most widely used neuropsychological measure, also yields strikingly large confidence intervals: in a study by Matarazzo and Herman (17), the 90% confidence interval for full-scale IQ in a group of 119 healthy comparison subjects had a 16.8-point difference. Using the reliable change index (corrected for practice) with 90% cutoff values, Sawrie and colleagues (23) found t intervals of 36 points for the WMS-R general memory index and 40 points for the WMS-R attention/concentration

index. Thus, large test-retest confidence intervals are not unique to the RBANS and illustrate the difficulty of assessing change at the level of an individual subject.

The confidence level or percentile chosen to aid in decision making should be based on the degree of confidence that the clinician feels is adequate for making treatment decisions. Such decisions, however, are rarely symmetrical, with practitioners most likely electing to err on the side of caution when faced with anticipated risks or evidence that suggests clinical deterioration. For example, to be outside of the 90% confidence limits, a conventional and conservative statistical standard, a patient's scores would need to show a retest decline greater than 13 points (we observed this level of difference in 4.4% of subjects). However, decreases of 8 or more points occurred in only 10% of subjects, and decreases of 6 or more points occurred in 14% of subjects. Thus, even relatively modest declines were infrequent. In light of this fact, a clinically conservative approach to decision making might warrant seeking additional corroborating evidence of deterioration when faced with declines in test scores that approach, but do not reach, conservative statistical standards. In the absence of such corroborating evidence, it is difficult to justify decision making that is based on a misunderstanding of the limits of measurement reliability (i.e., deciding that a patient has truly deteriorated clinically when a 6-point test-retest decline has been observed). Such weighting of clinical judgment versus statistical confidence varies as a function of the risks and costs of decision making.

Certainly, there are other means of establishing test stability. However, for the purposes of aiding a wide variety of clinical professionals to make treatment decisions, our method is straightforward and easily applied. More sophisticated multiple regression models would be able to narrow the confidence interval somewhat (24); however, these models pose the disadvantage of being less user friendly. Consequently, they might be less frequently applied in a treatment setting. Other methods have also been proposed (15, 22). However, they are inappropriate for tests with alternate forms and offer no substantial advantage in terms of reducing confidence intervals.

Overall, the RBANS performed with adequate test-retest stability, given that the instrument was designed as a brief screening tool and can be used by professionals with varying levels of experience with neuropsychological testing. Other test batteries might provide greater test-retest stability and additional measures of interest (e.g., the RBANS does not measure problem-solving/executive function). In accomplish these aims, however, test batteries would likely have to be substantially longer, which may be prohibitive in some settings. Moreover, the gains in reliability achieved by increasing test administration time may be surprisingly modest, given the apparent comparability of the RBANS and the Wechsler scales. Alternatively, the Mini-Mental State Examination (27), designed for its brevity of administration, does not appear adequately sensi-

tive to the level and pattern of impairment observed in many patients with schizophrenia (28). In addition, the Mini-Mental State Examination has shown less sensitivity to disease progression in Alzheimer's disease patients than the Mattis Dementia Rating Scale (29), a test that resembles the RBANS in terms of administration time. Thus, the RBANS appears to offer both the sensitivity and reliability necessary for repeated assessments. In addition, the RBANS offers age-scaled norms and an alternate form to aid in circumventing age and practice effects. Although it may have some limitations, the RBANS is a clinically useful screening instrument with test-retest stability that compares favorably with other commonly used psychometric instruments.

Received May 22, 2001; revision received Sept. 27, 2001; accepted Oct. 8, 2001. From the Maryland Psychiatric Research Center; Sheppard Pratt Health System, Baltimore; Chestnut Lodge Hospital, Rockville, Md.; Stanley Foundation Research Programs, Bethesda, Md.; and Loyola University Chicago. Address reprint requests to Dr. Gold, Maryland Psychiatric Research Center, P.O. Box 21247, Baltimore, Md. 21228-0747; jgold@mprc.umaryland.edu (e-mail).

Supported in part by NIMH grant MH-40279 and a grant from the Stanley Foundation.

References

1. Braff DL, Heaton RK, Kuck J, Cullum M, Moranville J, Grant I, Zisook S: The generalized pattern of neuropsychological deficits in outpatients with chronic schizophrenia with heterogeneous Wisconsin Card Sorting Test results. *Arch Gen Psychiatry* 1991; 48:891-898
2. Gold JM, Randolph C, Carpenter CJ, Goldberg TE, Weinberger DR: Forms of memory failure in schizophrenia. *J Abnorm Psychol* 1992; 101:487-494
3. Saykin AJ, Gur RC, Gur RE, Mozley PD, Mozley LH, Resnick SM, Kester DB, Stafiniak P: Neuropsychological function in schizophrenia: selective impairment in memory and learning. *Arch Gen Psychiatry* 1991; 48:618-624
4. Calev A: Neuropsychology of schizophrenia and related disorders, in *Assessment of Neuropsychological Functions in Psychiatric Disorders*. Edited by Calev A. Washington, DC, American Psychiatric Press, 1999, pp 33-66
5. Heinrichs RW, Zakzanis KK: Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* 1998; 12:426-445
6. Heaton R, Paulsen JS, McAdams LA, Kuck J, Zisook F, Braff D, Harris J, Jeste DV: Neuropsychological deficits in schizophrenia: relationship to age, chronicity, and dementia. *Arch Gen Psychiatry* 1994; 51:469-476
7. Saykin AJ, Shtasel DL, Gur RE, Kester DB, Mozley LH, Stafiniak P, Gur RC: Neuropsychological deficits in neuroleptic naive patients with first-episode schizophrenia. *Arch Gen Psychiatry* 1994; 51:124-131
8. Blyler CR, Gold JM: Cognitive effects of typical antipsychotic treatment: another look, in *Cognition in Schizophrenia: Impairments, Importance, and Treatment Strategies*. Edited by Sharma T, Harvey P. Oxford, UK, Oxford University Press, 2000, pp 241-265
9. Green MF: What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry* 1996; 153: 321-330
10. Randolph C: Repeatable Battery for the Assessment of Neuropsychological Status. San Antonio, Psychological Corp, 1998

11. Gold JM, Queern C, Iannone VN, Buchanan RW: Repeatable Battery for the Assessment of Neuropsychological Status as a screening test in schizophrenia, I: sensitivity, reliability, and validity. *Am J Psychiatry* 1999; 156:1937–1950
12. Hobart MP, Goldberg R, Bartko JJ, Gold JM: Repeatable Battery for the Assessment of Neuropsychological Status as a screening test in schizophrenia, II: convergent/discriminant validity and diagnostic group comparisons. *Am J Psychiatry* 1999; 156:1951–1957
13. McCaffrey RJ, Ortega A, Orsillo SM, Nelles WB, Haase RF: Practice effects in repeated neuropsychological assessments. *Clin Neuropsychol* 1992; 6:32–42
14. Bartko JJ, Carpenter WT: On the methods and theory of reliability. *J Nerv Ment Dis* 1976; 163:307–317
15. Chelune GJ, Naugle RI, Lueders H, Sedlak J, Awad IA: Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology* 1993; 7:41–52
16. Matarazzo JD, Carmody TP, Jacobs LD: Test-retest reliability of the WAIS: a literature review with implications for clinical practice. *J Clin Neuropsychol* 1980; 2:89–105
17. Matarazzo JD, Herman DO: Base rate data for the WAIS-R: test-retest stability and VIQ-PIQ differences. *J Clin Neuropsychol* 1984; 6:351–366
18. Wechsler D: Wechsler Adult Intelligence Scale—Revised. New York, Psychological Corp, 1981
19. Bruggemans EF, Van de Vijver FJR, Huysmans HA: Assessment of cognitive deterioration in individual patients following cardiac surgery: correcting for measurement error and practice effects. *J Clin Exp Neuropsychol* 1997; 19:543–559
20. Crawford JR, Howell DC: Regression equations in clinical neuropsychology: an evaluation of statistical methods for comparing predicted and obtained scores. *J Clin Exp Neuropsychol* 1998; 20:755–762
21. Dikmen SS, Heaton RK, Grant I, Temkin NR: Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *J Int Neuropsychol Soc* 1999; 5:346–356
22. Jacobson NS, Truax P: Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; 59:12–19
23. Sawrie SM, Chelune GJ, Naugle RI, Lueders HO: Empirical methods for assessing meaningful neuropsychological change following epilepsy surgery. *J Int Neuropsychol Soc* 1996; 2:556–564
24. Temkin NR, Heaton RK, Grant I, Dikmen SS: Detecting significant change in neuropsychological test performance: a comparison of four models. *J Int Neuropsychol Soc* 1999; 5:357–369
25. Fenton WS, Dickerson F, Boronow J, Hibbeln JR, Knable M: A placebo-controlled trial of omega-3 fatty acid (ethyl eicosapentaenoic acid) supplementation for residual symptoms and cognitive impairment in schizophrenia. *Am J Psychiatry* 2001; 158:2071–2074
26. WAIS-III/WMS-III Technical Manual. San Antonio, Tex, Psychological Corp (Harcourt), 1997
27. Folstein MF, Folstein SE, McHugh PR: “Mini-Mental State”: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12:189–198
28. Faustman MO, Moses JA, Csernansky JG: Limitations of the Mini-Mental State Examination in predicting neuropsychological functioning in a psychiatric sample. *Acta Psychiatr Scand* 1990; 81:126–131
29. Salmon DP, Thal LJ, Butters N, Heindel WC: Longitudinal evaluation of dementia of the Alzheimer type: a comparison of 3 standardized mental status examinations. *Neurology* 1990; 40:1225–1230