

## Test-Retest Reliability of a Functional MRI Working Memory Paradigm in Normal and Schizophrenic Subjects

Dara S. Manoach, Ph.D.

Elkan F. Halpern, Ph.D.

Todd S. Kramer, B.S.

Yuchiao Chang, Ph.D.

Donald C. Goff, M.D.

Scott L. Rauch, M.D.

David N. Kennedy, Ph.D.

Randy L. Gollub, M.D., Ph.D.

**Objective:** Repeated functional magnetic resonance imaging (fMRI) studies of schizophrenic subjects may identify brain activity changes in response to interventions. To interpret the findings, however, it is crucial to know the test-retest reliability of the measures used.

**Method:** The authors scanned seven normal subjects and seven schizophrenic subjects on two occasions during performance of a working memory task. They quantified the reliability of task performance and brain activation.

**Results:** In both groups, task performance was reliable, and all a priori regions were activated in group-averaged test and retest data. In individual schizophrenic subjects, however, indices of cognitive activation were not reliable across sessions. Normal subjects showed reasonable reliability of activation.

**Conclusions:** Even given reliable task performance, stable clinical status, and a stable pattern of group-averaged activation, individual subjects showed unreliable brain activation. This suggests that repeated fMRI studies of schizophrenia should control for sources of variation, both artifactual and intrinsic.

(*Am J Psychiatry* 2001; 158:955–958)

In schizophrenia, repeated functional magnetic resonance imaging (fMRI) studies may allow a determination of brain activity changes over time (e.g., over the course of illness, and in response to changes in symptom profile and to pharmacologic or other interventions). To evaluate the findings of repeated studies, however, it is crucial to know the test-retest reliability of the measures employed. We scanned normal and schizophrenic subjects on two occasions during working memory task performance using the Sternberg Item Recognition Paradigm (1) modified for fMRI (2). The Sternberg Item Recognition Paradigm is a continuous performance, choice reaction time task that has been demonstrated to be highly reliable and relatively free of practice effects in normal subjects (3). We quantified the reliability of Sternberg Item Recognition Paradigm performance and of associated activation in several a priori brain regions (4) that are thought to comprise the anatomical components of a working memory network (5). These include the dorsolateral prefrontal cortex, intraparietal sulcus, insula, supplementary motor area, lateral premotor, and primary motor areas.

### Method

A previous report (4) provides detailed descriptions of the methods employed. Seven normal and seven schizophrenic subjects participated (five men and two women in each group). Rescanning of schizophrenic subjects was contingent on no changes in medication and stable clinical status, as indicated by the Positive and Negative Syndrome Scale (6) (score within 20 points of the initial rating). Six normal and five schizophrenic subjects were right-handed. The interscan interval in weeks did

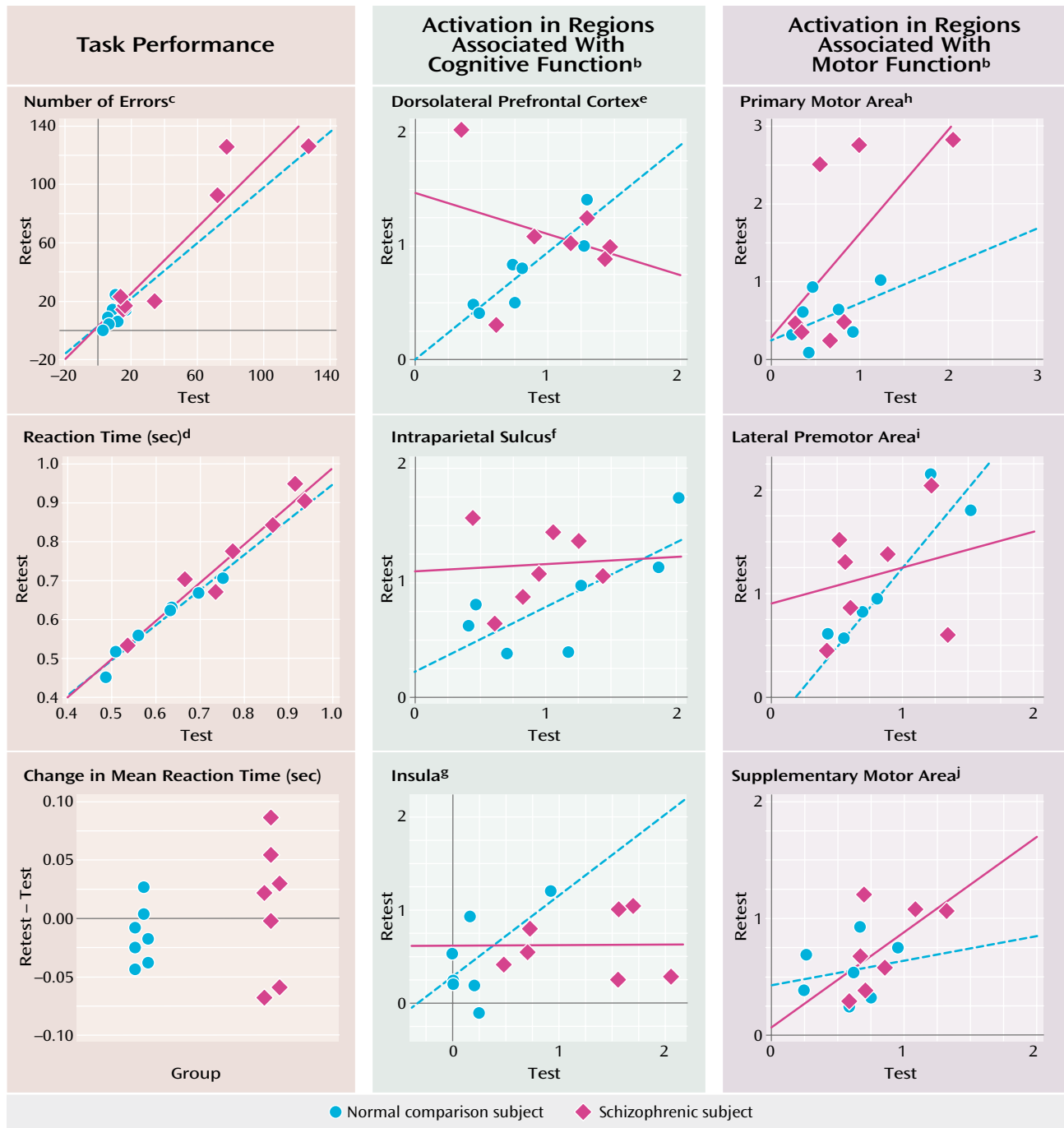
not differ significantly between groups (normal subjects: mean=13.6, SD=11.2; schizophrenic subjects: mean=9.6, SD=10.8). Normal subjects were younger (mean=34 years, SD=4, versus mean=45, SD=8) ( $t=3.59$ ,  $df=12$ ,  $p=0.004$ ) and had higher estimated verbal IQs (mean=126, SD=7, versus mean=102, SD=12) ( $t=5.28$ ,  $df=12$ ,  $p<0.001$ ). After a complete description of the study, all subjects gave written informed consent.

During working memory task performance, subjects were visually presented with either two or five digits to maintain in working memory. Subjects then responded to the presentation of single digits by pressing a button with either their right or left hand to indicate whether or not the digit was in the memorized set of digits. In the baseline condition, subjects responded to arrows pointing right or left by pressing the corresponding button. To maximize motivation, subjects were paid 5 cents for each correct response (total possible responses=396).

Functional images were collected with a General Electric (Milwaukee) Signa 1.5-T high-speed scanner by using a gradient echo T<sub>2</sub>-weighted sequence (TR=2000 msec, TE=50 msec, flip angle=70°). Fifteen contiguous interleaved axial slices (voxel size: 3.13 × 3.13 × 8 mm) were acquired. Functional scans were corrected for motion, normalized, vertically averaged for each subject, transformed into Talairach and Tournoux space (7), and vertically averaged for each group. We identified activated voxels using pairwise *t* tests of task conditions and a threshold of  $p<1 \times 10^{-4}$ . The dorsolateral prefrontal cortex was defined according to conservative Talairach and Tournoux coordinates (7). A neuroanatomist defined Talairach and Tournoux criteria for the other regions of interest. We measured the magnitude and spatial extent of activation in each region as the percent signal change in the voxel with the maximum *t* statistic and as the number of activated voxels.

We computed intraclass correlation coefficients (ICCs) to quantify the reliability of task performance and activation. ICCs represent the proportion of total variability accounted for by the variability between, rather than within, subjects. We considered ICCs of 0.50 or more to indicate reasonable reliability. We also computed regressions to determine whether the slope and intercept of the

FIGURE 1. Test and Retest Data for Seven Normal and Seven Schizophrenic Subjects During Performance of a Working Memory Task<sup>a</sup>



<sup>a</sup> Regression lines are presented in blue for the normal group and red for the schizophrenic group. Perfect test-retest agreement would be indicated by all points falling on a diagonal line beginning at the origin and bisecting the plot. Intraclass correlations (ICCs) were calculated to quantify reliability; values of 0.50 and higher indicate reasonable reliability.

<sup>b</sup> Percent signal change in the voxel with the maximum t statistic from the comparison of the five digits working memory condition to the arrows baseline condition.

<sup>c</sup> Comparison: ICC=0.44, schizophrenia: ICC=0.90.

<sup>d</sup> Comparison: ICC=0.93, schizophrenia: ICC=0.92.

<sup>e</sup> Comparison: ICC=0.81, schizophrenia: ICC=-0.20.

<sup>f</sup> Comparison: ICC=0.68, schizophrenia: ICC=0.07.

<sup>g</sup> Comparison: ICC=0.49, schizophrenia: ICC=0.01.

<sup>h</sup> Comparison: ICC=0.50, schizophrenia: ICC=0.46.

<sup>i</sup> Comparison: ICC=0.71, schizophrenia: ICC=0.17.

<sup>j</sup> Comparison: ICC=0.23, schizophrenia: ICC=0.57.

line describing the test-retest relationship differed from the agreement line. We present only the analyses of the five digits versus arrows (baseline condition) comparison, since it produced activation in the greatest number of subjects during the test session.

## Results

The schizophrenic subjects performed reliably with regard to the number of errors across conditions. Normal subjects performed near ceiling levels during both sessions, and their low ICC reflects this restricted range of errors (Figure 1, upper left). Mean reaction time across conditions was reliable for both groups (Figure 1, middle left). We also examined the variability of task performance. Although the difference in mean reaction time across sessions was close to zero and did not differ between groups ( $t=1.05$ ,  $df=12$ ,  $p=0.32$ ), the schizophrenic group tended to have greater variability in the magnitude and direction of change in reaction time across sessions ( $F=0.20$ ,  $df=1, 6$ ,  $p=0.07$ ) (Figure 1, lower left). The schizophrenic subjects also showed significantly greater within-subject variability of reaction time, as indicated by coefficients of variation ( $F=7.59$ ,  $df=1, 12$ ,  $p=0.02$ ).

In the group-averaged data, both groups activated all a priori regions at both sessions. However, when we examined activation from individual subjects as measured by the percent change in the voxel with the maximum  $t$  statistic, the schizophrenic subjects consistently showed less reliable activation than the normal subjects in regions associated with cognition (dorsolateral prefrontal cortex, intraparietal sulcus, and insula) (Figure 1, middle column). Their reliability was comparable to that of normal subjects in the primary motor area, worse in the lateral premotor area, and better in the supplementary motor area (Figure 1, last column) regions associated with motor function. In spite of the limited power of the study, regression lines were either significantly different or tended to differ from the agreement line only in the cognitive regions and only for the schizophrenic group. The pattern of findings using the number of activated voxels as the index of activation was similar.

## Discussion

Despite reliable task performance, stable clinical status, and a comparable group-averaged activation pattern, individual schizophrenic subjects showed unreliable activation in brain regions associated with cognitive function. Reliability was somewhat better in motor regions. The normal subjects, in contrast, showed reliable task performance and reasonable reliability of activation in four of the six regions studied. However, the normal subjects differed from the schizophrenic subjects in several important respects. They were younger, performed more accurately and less variably, had higher verbal IQs, and moved less during scanning. Although age and IQ differences limit the validity of the group comparison, all the schizophrenic subjects had at least average IQs, and the 11-year differ-

ence in mean age is unlikely to fully account for the differential reliability.

Greater motion and less accurate, more variable performance are fairly typical results in neuroimaging studies of schizophrenia. They frequently confound group comparisons and likely contribute to unreliable activation. Figure 1 reveals that several extreme cases (outliers) disproportionately contributed to unreliable cognitive activation in the schizophrenia group. Although one of these outliers made the most errors, the other two performed near the group mean for errors and variability, and none were outliers with regard to motion. Thus, the least reliable subjects were unlikely to be consistently identified by poor performance or greater motion.

We previously reported greater variability in the spatial distribution of activation *between* schizophrenic subjects during a single scanning session (4). The current study demonstrates unreliability of the magnitude and spatial extent of activation *within* schizophrenic subjects across scanning sessions. Greater variability may be intrinsic to schizophrenia. It has been documented even for simple manual preference tasks (8). Both aberrant performance and unreliable activation may reflect a failure to automatize task performance (e.g., to use practice to shape the optimal spatiotemporal pattern of activity in neural circuitry).

Although this is a small preliminary study, it has important implications for repeated fMRI studies in schizophrenia. The findings suggest that it is important to control for sources of variation (both artifactual and intrinsic), to have enough observations to account for greater variability, and to demonstrate reliability in schizophrenic subjects as well as in normal subjects.

---

Received July 18, 2000; revision received Dec. 29, 2000; accepted Jan. 29, 2001. From the Department of Neurology, Beth Israel Deaconess Medical Center; and the Departments of Psychiatry and Radiology, Medical Practices Evaluation Center, and the Center for Morphometric Analysis, Massachusetts General Hospital, Harvard Medical School, Boston. Address reprint requests to Dr. Manoach, Behavioral Neurology Unit, Beth Israel Deaconess Medical Center, 330 Brookline Ave., Boston, MA 02215; dmanoach@caregroup.harvard.edu (e-mail).

Funded by the Scottish Rite Benevolent Foundation's Schizophrenia Research Program, N.M.J., U.S.A. (to Dr. Manoach), the Mathers Charitable Foundation (to Dr. Manoach), and National Institute on Drug Abuse grants PA-00275 and PA-11229 (to Drs. Manoach and Gollub and Mr. Kramer).

The authors thank Meghan Searl, Ed Amico, Etienne Benson, and especially Nikos Makris, Ph.D., who defined the boundaries of the regions of interest.

---

## References

1. Sternberg S: High-speed scanning in human memory. *Science* 1966; 153:652-654
2. Manoach DS, Schlaug G, Siewert B, Darby DG, Bly BM, Benfield A, Edelman RR, Warach S: Prefrontal cortex fMRI signal changes are correlated with working memory load. *Neuroreport* 1997; 8: 545-549
3. Kristofferson MW: Effects of practice on character-classification performance. *Can J Psychol* 1972; 26:54-60

## BRIEF REPORTS

4. Manoach DS, Gollub RL, Benson ES, Searl MM, Goff DC, Halpern E, Saper CB, Rauch SL: Schizophrenic subjects show aberrant fMRI activation of dorsolateral prefrontal cortex and basal ganglia during working memory performance. *Biol Psychiatry* 2000; 48:99–109
5. Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE: Temporal dynamics of brain activation during a working memory task. *Nature* 1997; 386:604–608
6. Kay SR, Opler LA, Lindenmayer JP: The Positive and Negative Syndrome Scale (PANSS): rationale and standardisation. *Br J Psychiatry* 1989; 155:59–67
7. Talairach J, Tournoux P: *Co-Planar Stereotaxic Atlas of the Human Brain*. New York, Thieme Medical, 1988
8. Nelson LD, Satz P, Green M, Cicchetti D: Re-examining handedness in schizophrenia: now you see it—now you don't! *J Clin Exp Neuropsychol* 1993; 15:149–158