npg

Baseline Neurocognitive Deficits in the CATIE Schizophrenia Trial

Richard SE Keefe*,[1], Robert M Bilder[2,3], Philip D Harvey[4], Sonia M Davis[5], Barton W Palmer[6], James M Gold[7], Herbert Y Meltzer[8], Michael F Green[2], Del D Miller[9], Jose M Canive[10], Lawrence W Adler[11], Theo C Manschreck[12], Marvin Swartz[1], Robert Rosenheck[13], Diana O Perkins[14], Trina M Walker[15], T Scott Stroup[14], Joseph P McEvoy[1,15] and Jeffrey A Lieberman[16]

[1]Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA; [2]Department of Psychiatry & Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA; [3]Department of Psychology, Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, USA; [4]Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, USA; [5]Quintiles Inc., Morrisville, NC, USA; [6]Department of Psychiatry, University of California, San Diego, CA, USA; [7]Maryland Psychiatry Research Center, Baltimore, MD, USA; [8]Department of Psychiatry, Vanderbilt University Medical Center, Nashville, TN, USA; [9]Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA, USA; [10]Psychiatry Service, New Mexico VA Health Care System and Department of Psychiatry, University of New Mexico, Albuquerque, NM, USA; [11]Clinical Insights, Glen Burnie, MD, USA; [12]Corrigan Mental Health Center, Department of Psychiatry, Harvard Medical School, Boston, MA, USA; [13]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; [14]Department of Psychiatry, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; [15]Department of Biological Psychiatry, John Umstead Hospital, Duke University Medical Center, Durham, NC, USA; [16]Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York State Psychiatric Institute, New York, NY, USA

Neurocognition is moderately to severely impaired in patients with schizophrenia. However, the factor structure of the various neurocognitive deficits, the relationship with symptoms and other variables, and the minimum amount of testing required to determine an adequate composite score has not been determined in typical patients with schizophrenia. An 'all-comer' approach to cognition is needed, as provided by the baseline assessment of an unprecedented number of patients in the CATIE (Clinical Antipsychotic Trials of Intervention Effectiveness) schizophrenia trial. From academic sites and treatment providers representative of the community, 1493 patients with chronic schizophrenia were entered into the study, including those with medical comorbidity and substance abuse. Eleven neurocognitive tests were administered, resulting in 24 individual scores reduced to nine neurocognitive outcome measures, five domain scores and a composite score. Despite minimal screening procedures, 91.2% of patients provided meaningful neurocognitive data. Exploratory principal components analysis yielded one factor accounting for 45% of the test variance. Confirmatory factor analysis showed that a single-factor model comprised of five domain scores was the best fit. The correlations among the factors were medium to high, and scores on individual factors were very highly correlated with the single composite score. Neurocognitive deficits were modestly correlated with negative symptom severity ($r = 0.13–0.27$), but correlations with positive symptom severity were near zero ($r < 0.08$). Even in an 'all-comer' clinical trial, neurocognitive deficits can be assessed in the overwhelming majority of patients, and the severity of impairment is similar to meta-analytic estimates. Multiple analyses suggested that a broad cognitive deficit characterizes this sample. These deficits are modestly related to negative symptoms and essentially independent of positive symptom severity.
Neuropsychopharmacology (2006) **31,** 2033–2046. doi:10.1038/sj.npp.1301072; published online 19 April 2006

Keywords: cognition; neuropsychology; schizophrenia; clinical trials methodology; antipsychotics

---

INTRODUCTION

Neurocognition is moderately to severely impaired in patients with schizophrenia (Heinrichs and Zakzanis, 1998), and almost all patients demonstrate some measure of decline from their expected level of neurocognitive function (Keefe *et al*, 2005). Neurocognitive impairment is a core feature of schizophrenia, and not simply the result of the symptoms or the current treatments of the illness (Green *et al*, 2004b). Neurocognitive deficits are the single strongest correlate of real-world functioning (Green, 1996),

and are currently viewed as potential targets for treatment in patients with schizophrenia (Hyman and Fenton, 2003).

While the number of publications on neurocognitive deficits in schizophrenia has grown vastly over the past two decades (Hyman and Fenton, 2003), most reports have been generated from sample sizes that are considered small given the substantial heterogeneity of performance among patients with schizophrenia (Wilk et al, 2004). Furthermore, most studies of cognition have used samples of convenience that have included groups of patients who can complete extensive research protocols owing to their chronic institutionalization or who are available because they have entered a clinical trial or other specialized research study, and have thus been carefully screened with extensive exclusion criteria, such as the absence of substance abuse and medical comorbidities. These studies thus present a relatively narrow view of a filtered group of patients who may not reflect the profile of neurocognitive deficits, symptoms and substance abuse found in the overall population of patients with schizophrenia. Owing to these limitations, important questions about cognition in the majority of 'real-world' patients with schizophrenia are unknown: (1) What is the factor structure of neurocognitive performance in schizophrenia? (2) What is the relationship of neurocognitive deficits to symptoms and demographic factors? and (3) What is the amount of testing that is required to generate a meaningful estimate of neurocognitive function? Answers to these questions have profound implications for the methodology that will be used for future clinical trials of medications that may improve cognition in schizophrenia, including not only the neurocognitive measures that will be employed, but also what kinds of patients may enter these trials.

The CATIE (Clinical Antipsychotic Trials of Intervention Effectiveness) project was designed specifically to assess the treatment effectiveness of the available second generation antipsychotic medications and the conventional antipsychotic perphenazine in patients who are representative of the general population of people with schizophrenia—a so called 'all-comers' trial. This trial presents an ideal opportunity to assess the severity of neurocognitive performance and the relation of neurocognitive deficits to symptoms and demographic factors in patients who may better represent the typical patient with schizophrenia, including those who often do not participate in research protocols or clinical trials. As the sample size of this study ($N = 1460$) was large, it provides a unique opportunity to conduct analyses not possible with smaller samples: comparison of competing models of the structure of neurocognitive deficits in schizophrenia; determination of the magnitude of the correlations of neurocognitive deficits with symptom severity; and determination of the specific contributions of each test added to a battery for generating an estimate of overall neurocognitive functioning.

METHODS

The CATIE study compares the effectiveness of atypical medications and a single conventional antipsychotic medication through a randomized clinical trial involving a large sample of patients treated for schizophrenia at multiple sites, including both academic and more representative community providers. Participants gave written informed consent to participate in protocols approved by local IRBs. Details of the study design and entry criteria have been presented elsewhere (Stroup et al, 2003; Keefe et al, 2003). The current report relies exclusively on baseline data collected before randomization and the initiation of experimental treatments.

Patients

In total, 1493 patients with a diagnosis of schizophrenia, confirmed by the SCID (First et al, 1996), were entered into the study from 57 institutions. These sites directly reflect the US population in terms of the percentage from metropolitan areas with over 1 000 000 population (69% in CATIE; 73% in US). The sites were located in 48 cities and towns representing every region of the US except Alaska. Thus, the clinical sites are typical of the 'real world' settings in which people with schizophrenia receive treatment in the US (Swanson et al, in press). Sites with 15 or fewer patients were combined into pooled sites based on the type of site, and data from one site (33 patients) were excluded from all analyses because of concerns about data integrity, yielding 49 total sites. In Table 1, the clinical and demographic characteristics of the patients in the study are presented for the entire sample ($N = 1460$) and for four different subgroups: those who completed all 11 tests ($N = 1005$), those with 1–3 missing tests ($N = 327$), those with more than three missing tests ($N = 82$), and those who were not tested at all ($N = 46$). Given these large sample sizes and the associated statistical power, the emphasis of the results will focus, when possible, on effect sizes and z-scores. There were subtle differences between groups on age, education, years since first antipsychotic treatment, but no differences between groups on PANSS scores and all other demographic variables. Patients who completed all of the neurocognitive tests tended to be younger than the other groups, have more education, and have a shorter period of time since their first antipsychotic treatment. All of these differences, however, were less than 0.4 standard deviations.

Neurocognitive Measures

Table 2 describes the tests, neurocognitive domains, and measures collected in the CATIE battery. All tests and the rationale for their use, including their reliability and validity, as well as the procedures for training testers, are described in greater detail in a previous publication on the methodology for this study (Keefe et al, 2003). To assess the effect of previous experience on test performance, all patients were asked whether they had ever received each of the tests.

Clinical Assessments

Clinical raters, whose inter-rater reliability was established in a group setting, determined scores on the Clinical Global Impression (CGI; Guy, 1976) and the Positive and Negative Syndrome Scale (PANSS; Kay et al, 1987). The CGI measured the overall severity of respondents' illness on a 7-point scale from 1 = normal, not ill to 7 = very severely ill.

Table 1 Demographic Characteristics of all CATIE Patients and Patients with Various Levels of Neurocognitive Data

Variable	All patients (N = 1460) Mean	SD	Complete data (N = 1005) Mean	SD	Missing ⩽ 3 tests (N = 327) Mean	SD	Missing > 3 tests (N = 82) Mean	SD	No data (N = 46) Mean	SD	Analysis F^a	p
Age (years)	40.56	11.10	39.58	10.92	43.04	10.82	41.56	12.66	42.41	11.38	8.83	<0.0001
Patient's education (years)	12.10	2.25	12.28	2.09	11.68	2.52	11.91	2.43	11.59	2.94	6.81	<0.0001
Duration since first prescribed antipsychotic medication (years)	14.43	10.72	13.48	10.28	17.10	11.32	14.93	12.18	15.37	10.19	9.24	<0.0001
PANSS (total score)	75.67	17.56	75.22	17.66	75.24	16.71	78.44	20.41	81.24	21.36	2.17	0.0895

	% Frequency	% Frequency	% Frequency	% Frequency	% Frequency	χ^2	p
Sex						7.6412	0.0540
Male	74	75	76	66	61	df = 3	
Female	26	25	24	34	39		
Race						6.2080	0.1019
White	60	61	60	51	48	df = 3	
Other	40	39	40	49	52		
Ethnic origin						3.9281	0.2693
Hispanic	12	11	13	16	17	df = 3	
Non-hispanic	88	89	87	84	83		
Marital status						10.4758	0.5743
Married	11	11	15	7	9	df = 12	
Widowed	2	2	3	4	2		
Divorced	21	21	20	16	22		
Separated	6	6	6	7	2		
Never married	60	60	57	66	65		
Employment status						6.0134	0.4217
Full time	7	7	8	4	3	df = 6	
Part time	8	9	7	5	8		
Unemployed	85	84	86	91	89		

[a]Statistical test for overall differences between four groups using ANOVA or a χ^2 test.
PANSS = positive and negative syndrome scale.

Table 2 Neurocognitive Tests Used in the CATIE Schizophrenia Trial

Cognitive domain	Type of test	Description of test and measures used for domain summary scores
Processing speed	1. Controlled oral word association test (Benton and Hamscher, 1978)	Generate as many words as possible beginning with F, A, and S in three separate trials of 60 s. *Measure:* mean number of words
	2. Category instances (Benton and Hamscher, 1978)	Name as many words as possible in 60 s within each of three categories (animals, fruits, vegetables). *Measure:* mean number of words
	3. Grooved pegboard (Lafayette Instrument Company, 1989)	Insert pegs in a certain order into 25 randomly positioned slots using dominant hand. Insert as many pegs as possible in two 45-s trials. *Measure:* mean number of pegs
	4. Wechsler adult intelligence scale-revised, digit symbol test (Wechsler, 1974)	Each numeral (1 through 9) is associated with a different simple symbol. Copy as many symbols associated with the numerals as possible in 90 s. *Measure:* number of symbols correctly copied
Reasoning	1. Wisconsin card sorting test, 64-card computerized version (Kongs *et al*, 2000)	A complex task of categorization, set shifting, and response to feedback from the tester. Only 64 cards were presented at each testing session. *Measure:* Mean of perseverative errors (sign reversed) and categories z-scores
	2. Wechsler Intelligence Scale for Children, third edition mazes (Wechsler, 1991)	Use a pencil to attempt to draw through a series of nine timed mazes without entering into blind alleys. *Measure:* raw score
Verbal memory	Hopkins verbal learning test (Brandt and Benedict, 1991)	Recall as many words as possible from a list of 12 words read aloud by the tester. The procedure was repeated three times. *Measure:* total number of words recalled
Working memory	1. Computerized test of visuospatial working memory (Lyons-Warren *et al*, 2004)	Focus on cross in center of computer screen while dot appears at random locations on the screen. Then watch shapes appear on screen and press space bar when diamond shape appears. When the cross reappears, point to where dot appeared in beginning. Distance between response and dot is measured. There are three conditions (no delay, 5-s delay and 15-s delay) and eight trials of each condition. *Measure:* Mean error distance of delay conditions minus no delay error distance (sign reversed).
	2. Letter–number sequencing test (Gold *et al*, 1997)	Clusters of letters combined with numbers (eg N6G2) are read aloud by the tester, and the patient reorganizes the cluster so that the numbers are listed first, from lowest to highest, followed by the letters in alphabetical order. *Measure:* number of correct sequences
Vigilance	Continuous performance test (CPT) (Cornblatt and Keilp, 1994)	Respond to a series of numbers on a computer screen at a rate of one per second by lifting a finger from a mouse key whenever the current number is identical to the previous number. Three 150-trial conditions of increasing difficulty are administered (2-, 3-, and 4-digit numbers, respectively). *Measure:* mean response sensitivity (D-prime).
Social cognition	Facial emotion discrimination test (Kerr and Neale, 1993)	Two pictures of different human faces displaying emotions are presented concurrently. Patients declare whether the faces are displaying the same or different emotions in thirty trials. *Measure:* number of correct responses.

The PANSS includes three scales and 30 items: seven items make up the Positive scale (examples are delusions, conceptual disorganization, and hallucinatory behavior), seven items make up the Negative scale (examples are blunted affect, emotional withdrawal, poor rapport, and passive/apathetic social withdrawal), and 16 items make up the General Psychopathology scale (examples are somatic concern, anxiety, guilt feelings, mannerisms and posturing, motor retardation, uncooperativeness, disorientation, poor impulse control, and preoccupation). PANSS ratings were made after the completion of a semistructured interview plus additional reports of daily function from caregivers or family members and a review of available clinical information. Individual items were scored on a Likert scale; values range from 1 to 7. Scores of 2 indicated that the presence of a symptom was questionable; scores above 2 indicated that a clinical symptom is present, and ratings of 3–7 indicated increasing severity. Detailed anchors were provided for severity ratings.

Illness duration was calculated as the number of years since the first antipsychotic treatment. Education level was collected in nine categories from 'did not complete high school' to 'advanced degree completed,' and was converted to approximate years of education.

Data Cleaning, Reduction, and Analytic Plan

Statistical characteristics of measures. For each collected and calculated measure, the mean, standard deviation, median, mode, skewness, and kurtosis were determined. All variables except the Facial Emotion Discrimination Task were judged to have reasonable normality such that

subsequent analyses could proceed. While three of the individual measures that were used to derive cognitive outcome measures (eg the no delay visuospatial working memory test performance) were skewed, all of the measures that comprised the domain scores (described below), the domain scores themselves, and the composite scores (described below) had skewness indices of between 0.45 and −0.25, suggesting considerably normality. Scores that were judged to reflect a patient's inability to complete the task were set to 'missing' in the data set; some scores generated queries to the tester to assure validity (see Appendix II for a description of scores that were queried or set to 'missing'). The data collected from the Facial Emotion Discrimination Test had a ceiling effect and positive skew that was so severe that it would be likely to reduce the sensitivity of the battery to treatment effects, so it was eliminated from further analyses.

Domain summary scores. All test measures were first converted to standardized z-scores by setting the mean of each measure to 0 and the standard deviation to 1 based upon the entire patient sample. The controlled oral word association test and category instances summary measures were averaged together to form one summary test score referred to as verbal fluency. For domains that had more than one measure, summary scores were determined by calculating the mean of the z-scores for the measures that comprised the domain (as described in Table 2), then converting the mean to a z-score with mean of 0 and standard deviation of 1. This process resulted in nine test summary scores and five domain scores (after excluding facial emotion discrimination) that correspond to five of the seven domains in the MATRICS consensus battery (Green *et al*, 2004b).

Comparison of neurocognitive measures to published norms. For selected measures, the performances of the patients in this sample were compared to available published data on healthy controls as estimates of the general population, commonly referred to as 'normative data'. The choice of which measures had sufficient normative data was based upon these characteristics, listed in order of importance: sample size; availability of education or IQ data; similarity of the mean age of the normative sample to the CATIE sample; and availability of other demographic data such as race and sex.

Data structure and calculation of composite score. Data were missing for 16.9% of the patients for one of the 11 measures, 3.5% for two measures, 2.9% for 3 measures, and an additional 0.6% for four of the 11 measures. This distribution suggested a natural nodal point between three and four measures, and three measures was chosen as a criterion for the maximum number of measures that could be missing if a patient were to be included in analyses using a composite score. Therefore, all composite score calculations and analyses are limited to the 1332 patients with three or fewer missing test scores. A domain-based composite score was defined as the simple average of the five domain summary scores.

Principal components analysis (PCA) of the nine standardized test scores was performed to determine the structure of components in the neurocognitive data set. The PCA identified components with eigenvalues greater than 1. Composite scores were calculated in two different ways: a domain-based approach and an empirical approach using PCA. The domain-based composite score was the standardized average of the five domain summary scores, following the MATRICS 7-factor model (Nuechterlein *et al*, 2004) with the exception of visual memory, for which there was no assessment, and social cognition, for which the data were unacceptably skewed. The PCA composite score was created by applying the weights generated from the PCA to the standardized variables and then standardizing the mean.

The potential factor structure of the neurocognitive battery was examined by confirmatory factor analysis (CFA) via structural equation modeling with the AMOS program (Arbuckle, 2003). Four different substantive models were tested: a null model positing no intercorrelations of any neurocognitive measure, a unifactorial model positing that all nine variables loaded onto a single factor, a unifactorial model positing that the five predefined domain scores described in Table 2 loaded onto a single factor, and a five-factor model that posited a structure based upon the five neurocognitive domains.

The capacity to calculate a meaningful composite score for the largest number of patients was determined by calculating a series of unweighted averages of the test scores based only on people with complete data for the tests, while sequentially reducing the number of tests that comprise the average, beginning with the tests that were completed by the least number of patients. The effect of each sequential exclusion of a test was evaluated by the increase in sample size, and decrease in the Pearson correlation between the unweighted average of the included items, and the domain-based composite score.

The contribution of each test to an unweighted average of standardized test scores was calculated by multiple regression analyses entering each variable hierarchically (ie on the basis of how much variance in the composite score the variable accounted for), as well as a stepwise regression with variable entry based upon the length of time required to administer a test.

Inter-relationship of neurocognitive domains. The relationship of the neurocognitive domain scores with each other and with the composite score was determined with Pearson correlations. Cronbach's alpha was calculated to determine the effect of removing each domain from the composite score. In addition, correlations between each domain and the composite score with that domain removed were calculated

Relationship to clinical and demographic factors. The relationship of the neurocognitive domain scores and the composite score to continuous demographic variables, CGI severity and PANSS total and subscale scores was calculated using Pearson correlations. Differences in the composite score between levels of categorical demographic variables was calculated with ANOVA.

Site effects. The amount of variance in the measures accounted for by site effects was calculated using ANOVA, with pooled site (N = 49) as the predictor variable. Sites

with 15 or fewer patients were combined into pooled sites based on site care setting.

Test experience. The effect of previous experience on performance was calculated for each test by dichotomizing patients into those who did and did not remember taking the test previously, and comparing the groups with ANOVA.

RESULTS

Table 3 presents raw score performance data for each collected and calculated neurocognitive measure. The table also includes comparison data from normative samples for measures with normative data judged to be of sufficient size and quality.

Data Structure

The PCA suggested that a single component best described the data, as the eigenvalue of the first principal component was 4.07, accounting for 45.2% of the overall variance in test scores, with no other principal component greater than 1.0. As this one component accounted for a relatively low percentage of the overall variance, 2-, 3-, and 4-component solutions were examined. These extractions followed by Varimax rotations, which emphasize differences among components, did not adequately explain the data structure.

In structural equation modeling analyses, the null model failed to fit the data, as indexed by a large and significant χ^2-statistic: $\chi^2(36) = 2861.1$, $p < 0.001$. A unifactorial model including the nine individual tests was a considerable improvement in fit, with a smaller χ^2, $\chi^2(27) = 192.18$,

Table 3 Neurocognitive Performance in Patients with Schizophrenia and Comparison to Available Normative Data

| Variable | Schizophrenia patients | | | Normative data | | |
	N	Mean	SD	Mean	SD	Group difference in mean[a]
WRAT-3 reading	1294	43.10	8.49	49.0	7.00	−0.84
COWAT—F words	1331	9.88	4.06			
COWAT—A words	1331	7.95	3.66			
COWAT—S words	1331	10.08	4.28			
COWAT sum	1331	27.91	10.63	38.93	12.17	−0.91
Category instances (animals)	1332	13.91	4.88			
Category instances (fruits)	1331	10.14	3.38			
Category instances (vegetables)	1332	8.64	3.44			
Category instances sum	1331	32.69	9.95			
WISC-III mazes	1319	17.54	5.98			
Letter–number sequencing	1323	10.21	4.47			
Hopkins verbal learning test trial 1	1331	4.75	1.74	7.63	1.77[b]	−1.63
Hopkins verbal learning test trial 2	1331	6.48	2.14	9.79	1.67[b]	−1.98
Hopkins verbal learning test trial 3	1331	7.50	2.45	10.74	1.31[b]	−2.47
Facial emotion discrimination test	1331	24.59	3.37			
WAIS-R digit symbol	1327	37.43	13.41	53.50	14.00	−1.15
Grooved pegboard trial 1	1323	11.75	3.93	15.36	4.24	−0.85
Grooved pegboard trial 2	1323	13.30	4.02			
CPT d', 2 digit	1095	2.37	1.05			
CPT d', 3 digit	1099	1.79	0.93			
CPT d', 4 digit	1081	1.01	0.76			
Visuospatial Working memory test no delay condition	1211	2.31	2.93			
Visuospatial working memory test 5 s condition	1211	28.28	19.08			
Visuospatial working memory test 15 s condition	1211	30.46	19.11			
Visuospatial working memory test (mean 5 and 15 s) minus no delay	1211	27.06	17.59			
Wisconsin card sorting test-64, perseverative errors	1263	13.66	10.18	8.50	6.30	0.82
Wisconsin card sorting test-64, categories completed	1260	2.13	1.60	3.56	1.47	−0.97
Wisconsin card sorting test-64, categories completed+additional cards in final category	1261	2.33	1.69			

COWAT = controlled oral word association test; WISC-III = Wechsler intelligence test for children, third edition; WAIS-R = Wechsler adult intelligence test, revised edition; CPT = continuous performance test, identical pairs.
[a]Based upon normative data z-score; all between-group comparisons were statistically significant at least $p < 0.0001$.
[b]SDs were estimated from Table 2 in the HVLT manual (Brandt and Benedict, 1991). Four age groups were pooled resulting in an age range of 20–60 years. The normative standard deviation was estimated by squaring individual SDs from each age group, dividing them by N-1 for that group, summing them, dividing by total N and taking the square root.

$p < 0.001$. Fit statistics were also reasonable: CFI = 0.94, GFI = 0.97, RMSEA = 0.077. This model was a significantly better fit than the null model, $\chi^2(9) = 2668.9$, $p < 0.001$. A unifactorial model including the five predefined domain scores was a considerable improvement in fit over the unifactorial model from the nine tests, with a smaller χ^2, $\chi^2(5) = 39.91$, $p < 0.001$, but similar fit statistics: CFI = 0.98, GFI = 0.97; RMSEA = 0.080. This model was a significantly better fit than the unifactorial model derived from the nine test scores, $\chi^2(22) = 152.27$, $p < 0.001$. A five-factor model that included the tests from each of the five cognitive domains as separate factors also fit the data reasonably well, $\chi^2(19) = 117.95$, $p < 0.001$, CFI = 0.96, GFI = 0.97, RMSEA = 0.071, and improved on the fit of the null model, $\chi^2(17) = 2734$, $p < 0.001$ and the unifactorial model from the nine tests, $\chi^2(8) = 74.23$, $p < 0.001$. However, it was a significantly poorer fit compared to the unifactorial model from the five predefined domain scores $\chi^2(14) = 78.04$, $p < 0.001$. None of the models fit the data well enough to meet traditional standards of a 'close fit' (RMSEA < 0.05) (Browne and Cudeck, 1993).

The structural consistency of the composite score was high, with Cronbach's alpha ranging from 0.077 to 0.82 if any of the five individual domain scores was removed from the composite score. Each domain had a very large correlation with the composite score (all r's < 0.72). Correlations between each domain score and a composite score that excluded that domain were as follows: verbal memory: $r = 0.55$; vigilance: $r = 0.56$; processing speed: $r = 0.74$; reasoning and problem solving: $r = 0.56$; and working memory: $r = 0.71$. There were also medium to large correlations among the five domain scores (minimum $r = 0.36$, maximum $r = 0.62$).

Statistical Characteristics of the Composite Score

The domain-based composite score and PCA composite score were highly correlated ($r = 0.988$, df = 1034, $p < 0.0001$), indicating that these scores have only 2.46% nonoverlap in variance. As a composite of the five domains, resulting in a single global score, had the best fit and appears to balance the competing requirements of following the theory-based factor structure, and parsimony in interpretation of the data, the domain-based composite score was chosen for all subsequent analyses. It should be noted, however, that this choice ultimately had negligible impact given that different approaches to computing the composite score yielded indices that are statistically interchangeable.

The effect of increasing the domain-based composite score sample size by sequentially reducing the number of tests that comprise it is presented in Table 4. Inspection of the tradeoff between increased sample size and reduction in correlation between the unweighted mean of the non-excluded tests and the domain-based composite score suggests a significant increase in sample size (from 1035 to 1297) if the three computerized tests are eliminated from the unweighted mean, with diminished subsequent gains in sample size if additional tests are eliminated. The Pearson correlation between the composite score and an unweighted mean generated from all of the noncomputerized tests was 0.951 (df = 1295, $p < 0.0001$), suggesting that the unweighted mean score created from all but these three tests contained

Table 4 Change in Pearson Correlation and Sample Size when Tests are Sequentially Removed from the Neurocognitive Composite Score in Order of Missing Case Size

Test removed from summary score	Correlation with composite score (r)	N (% of total)
None (complete data)	0.987	1035 (77.8)
Visuospatial working		
Memory test	0.980	1123 (84.3)
CPT	0.961	1230 (92.3)
WCST	0.951	1297 (97.4)
Pegboard	0.947	1306 (98.0)
WISC-III mazes	0.919	1319 (99.0)
HVLT total recall	0.886	1319 (99.0)
Letter–number sequencing	0.809	1327 (99.6)
WAIS-R digit symbol	0.658	1332 (100)
(Verbal fluency)[a]	(0.749)	(1327) (99.6)

[a]Since only five fewer patients had available data for WAIS-R digit symbol compared to verbal fluency, and since WAIS-R digit symbol accounted for more variance in the summary score than any other measure, we also report (in parentheses) the results of deleting all tests except WAIS-R digit symbol. CPT = continuous performance test, identical pairs; WCST = Wisconsin card sorting test; WISC-III = Wechsler intelligence test for children, third edition; HVLT = Hopkins verbal learning test; WAIS-R = Wechsler adult intelligence test, revised edition.

90.4% of the information from the domain-based composite score created from all of the tests. This result confirmed the use of three missing tests as a criterion for the maximum number of tests that could be missing if a patient were to be included in analyses using a composite score.

A summary score that includes only measures of WISC mazes, verbal fluency, HVLT total recall, WAIS digit symbol, and letter–number sequencing had a correlation of $r = 0.92$ with the composite score developed from all tests, yet results in missing data for only 13 of the 1332 patients who had three or fewer missing tests. An investigation of the 'comments' section of the database suggested that the most frequent reasons for missing data were 'out of range score' followed by 'computer malfunction' and 'patient refused test'.

The results of a stepwise multiple regression predicting an unweighted mean of all standardized test scores with variable entry based upon the estimated amount of time that a test takes to administer are presented in Table 5. This analysis includes only those patients who had data available for all of the nine measures included in the composite score ($N = 1035$). A hierarchical regression predicting the unweighted mean of all standardized test scores score with order of entry based upon the amount of variance that each variable accounted for is presented in Table 6. The most striking aspect of these analyses is that the overwhelming majority of the variance in the composite score was accounted for by a few tests.

Correlations Among Neurocognitive Measures

The Pearson correlations among the nine measures are presented in Table 7. The Wisconsin card sorting test

(WCST) score was the least strongly correlated with the other measures, with correlations ranging between 0.263 and 0.370. With one exception, all of the correlations among the other measures ranged between 0.317 and 0.556. Correlations with the composite score ranged between 0.543 and 0.749.

Site Effects

The amount of variance accounted for by the effect of site ranged between 6.3 (WCST) and 18.0% (WISC-III Mazes), with an average of 10.4% for all of the measures. Site effects accounted for 13.0% of the variance in the composite score. To determine if the site effects were due to site differences in education, the analyses were repeated using WRAT Reading score as a covariate measure. This covariance analysis did not reduce the magnitude of the site differences.

The Effect of Experience on Performance

The percentage of patients reporting that they had taken a specific test in the past ranged between 13.1% for the verbal fluency tests and 5.4% for the computerized visuospatial working memory test. Patients who had had experience with WAIS-R digit symbol performed significantly better on this test than those who did not have experience ($t = 2.88$, $df = 1306$, $p = 0.0041$); however, this difference ($d = 0.24$) was similar to Cohen's (1977) definition of a 'small' effect: $d = 0.2$. For all other tests, those patients with experience did not perform significantly different from those without experience.

Correlations with Demographic and Clinical Characteristics

The Pearson correlations of the neurocognitive domain scores and the composite score with demographic and clinical characteristics, including symptoms, are described

Table 5 Stepwise Multiple Regression Predicting Unweighted Mean of Variables from Each Test

Variable entered	Total R^2	R^2 Change	Estimated test admin time (mins)
WAIS-R digit symbol	0.610	0.610	3.4
HVLT total recall	0.722	0.112	4.1
Grooved pegboard	0.790	0.068	5.0
Letter–number sequencing	0.867	0.078	5.9
Verbal fluency	0.888	0.021	8.0
WISC-R mazes	0.934	0.046	11.2
Continuous performance test	0.956	0.022	13.4
Visuospatial working memory test	0.978	0.022	16.2
WCST	1.000	0.022	20.0

Order of entry from least to most estimated test administration time. F-statistic for all steps was greater than 193.0; all p-values < 0.0001; $N = 1035$. WAIS-R = Wechsler adult intelligence test, revised edition; HVLT = Hopkins verbal learning test; WISC-III = Wechsler intelligence test for children, third edition; WCST = Wisconsin card sorting test.

Table 6 Hierarchical Multiple Regression Predicting Unweighted Mean of Variables from Each Test

Variable entered	Total R^2	R^2 change
WAIS-R digit symbol	0.610	0.610
Letter–number sequencing	0.771	0.161
WISC-R mazes	0.847	0.076
Visuospatial working memory test	0.887	0.040
WCST	0.921	0.033
Verbal fluency	0.947	0.026
Grooved pegboard	0.968	0.021
HVLT total recall	0.984	0.016
CPT	1.000	0.016

Order of entry based upon amount of variance accounted for. F-statistic for all steps was greater than 364.0; all p-values < 0.0001; $N = 1035$. WAIS-R = Wechsler adult intelligence test, revised edition; WISC-III = Wechsler intelligence test for children, third edition; WCST = Wisconsin card sorting test; HVLT = Hopkins verbal learning test; CPT = continuous performance test, identical pairs.

Table 7 Correlation Matrix Among CATIE Neurocognitive Measures

Measure	Verbal fluency	Grooved pegboard	WAIS-R digit symbol	WCST	WISC-III mazes	HVLT verbal memory	VSWM deficit	LN span
Grooved pegboard	0.324							
WAIS-R digit symbol	0.528	0.546						
WCST	0.283	0.325	0.317					
WISC-III mazes	0.360	0.439	0.447	0.341				
HVLT total recall	0.508	0.289	0.466	0.273	0.325			
Visuospatial WM	0.326	0.380	0.446	0.263	0.410	0.381		
Letter–number span	0.492	0.372	0.521	0.370	0.412	0.528	0.437	
CPT	0.439	0.382	0.556	0.299	0.356	0.371	0.374	0.510
Composite	0.658	0.604	0.749	0.543	0.648	0.735	0.658	0.766

$p < 0.0001$ for all correlations. WAIS-R = Wechsler adult intelligence test, revised edition; WCST = Wisconsin card sorting test; WISC-III = Wechsler intelligence test for children, third edition; VSWM = visuospatial working memory; HVLT = Hopkins verbal learning test; LN = letter–number; CPT = continuous performance test, identical pairs.

in Table 8. The correlations varied very little between neurocognitive domains. The magnitude of the correlations with age, education, and duration of illness were in the medium range, defined by Cohen (1977) as $r = 0.30$, whereas the other correlations were smaller. The correlations with negative symptoms were in range defined by Cohen as small ($r = 0.10$) to medium ($r = 0.30$). Of particular note, all of the correlations between the neurocognitive measures and PANSS positive symptoms were nonsignificant and less than $r = 0.08$.

The following characteristics significantly differentiated patients' composite scores: race ($F = 25.65$, $df = 1,1328$, $p < 0.0001$), with white patients performing 0.28 standard deviations better than non-white patients; employment ($F = 13.77$, $df = 2$, 1316, $p < 0.0001$), with full-time and part-time employed patients performing, respectively, 0.49 and 0.30 standard deviations better than unemployed patients; and antipsychotic medication status at baseline ($F = 8.90$, $df = 2$, 1274, $p < 0.001$), with patients on no antipsychotic medication or atypicals performing similarly, and those on typicals at baseline performing about 0.32 standard deviations below the other two groups. Composite scores did not significantly differ by sex or whether a patient had been hospitalized or had an acute exacerbation in the past 3 months.

DISCUSSION

As this study was designed to accept all patients with chronic schizophrenia, and the majority of the patients were able to complete the cognitive battery, a broader range of patients was included in this study compared to previous work involving far smaller samples. The demographic and clinical characteristics of these patients were very similar to descriptions of patients receiving clinical treatment in the US (Slade et al, 2005; Swanson et al, in press).

Estimates of the severity of neurocognitive deficits in schizophrenia have relied upon selective synthesis of individual studies or meta-analyses, and have varied widely.

In the largest meta-analysis of schizophrenia patients published to date, patients performed between 0.46 and 1.41 standard deviations below controls on a range of 22 neuropsychological tests (Heinrichs and Zakzanis, 1998). However, individual estimates of the magnitude of severity of deficits have been variable, ranging up to 2 or 3 standard deviations below the healthy control mean (Harvey and Keefe 1997; Saykin et al, 1991). As this study was limited by the absence of a healthy control group, the severity of neurocognitive impairment in the patients in the current study was determined by comparing their performance to that of healthy controls for tests that had adequate published normative samples. The magnitude of impairment was similar to estimates from meta-analyses (Heinrichs and Zakzanis, 1998), large clinical trials (Harvey et al, 2003, 2004), and results obtained at academic research institutions (Keefe et al, 2004; Bilder et al, 2000; Heaton et al, 2001).

Although the relationship between substance abuse and neurocognition in schizophrenia is unclear (Green et al, 2004a; Bowie et al, 2005; Stirling et al, 2005), the inclusion of patients with substance abuse in this study may have increased the severity estimate compared to more purified samples. The effects of substance abuse on neurocognition and other clinical factors in this cohort will be reported in a separate manuscript. However, this comorbidity is a frequent occurrence in a large percentage of patients with schizophrenia, and the inclusion of these patients may yield a more realistic estimate of the severity of neurocognitive deficits in schizophrenia.

A PCA of the neurocognitive data from this study suggested that the battery yields a single principal component with all measures showing intercorrelations in the medium to large range. CFA suggested that a single-factor model of five domain scores, designed to estimate the domain scores following the model forwarded by the MATRICS Neurocognition Committee (Green et al, 2004b) based on a review of factor analytic studies (Nuechterlein et al, 2004), provided the best fit. A five-factor model also fit the data. However, the individual domains showed high

Table 8 Correlation Matrix of CATIE Neurocognitive Domains and Composite Score with Clinical and Demographic Factors

Domain	Theoretical composite (N = 1332)	Verbal memory (N = 1332)	Vigilance (N = 1212)	Processing speed (N = 1332)	Reasoning and problem solving (N = 1331)	Working memory (N = 1331)
Age	−0.305	−0.112	−0.176	−0.345	−0.363	−0.211
Education (years)	0.286	0.235	0.207	0.251	0.159	0.248
Duration of illness	−0.317	−0.139	−0.206	−0.331	−0.324	−0.246
WRAT-3 reading	0.471	0.322	0.373	0.408	0.306	0.412
CGI	−0.125	−0.124	−0.064	−0.072	−0.053	−0.147
Total PANSS	−0.166	−0.162	−0.108	−0.126	−0.074	−0.159
PANSS positive	−0.031	−0.054	0.005	0.025	−0.007	−0.073
PANSS negative	−0.271	−0.235	−0.196	−0.268	−0.133	−0.209
PANSS general	−0.108	−0.111	−0.073	−0.069	−0.043	−0.112

The N for the correlations varied based upon the neurocognitive measures. The symptom and demographic measures had very few missing cases, and are thus not reported for each correlation. All correlations greater than 0.10 are significant at $p < 0.0001$.
Duration of illness defined as number of years since first prescribed antipsychotic medication; WRAT-3 = wide range achievement test, third edition; CGI = clinical global impression; PANSS = positive and negative syndrome scale.

correlations among each other and with the composite score, even when the domain being correlated was removed from the composite scores. Reliability statistics suggested that the removal of any of the domain scores from the composite had little effect on the reliability of the composite score. These analyses suggest that the domains in this battery were very limited in their independence from one another. It remains unknown whether alternate test batteries might better identify discrete neurocognitive constructs impaired in schizophrenia.

This interpretation is limited by two important factors. First, the small number of measures (nine) entering the PCA and structural models limits the ability of these analyses to detect a more complex factorial structure of neurocognitive function. Support of more complex models will probably require a larger number of variables per domain, as demonstrated in the six-factor model supported in healthy subjects (Tulsky and Price, 2003) and schizophrenia patients (Nuechterlein et al, 2004; Gladsjo et al, 2004). However, adequate specification of independent cognitive factors is likely to require substantial additional measurement. Some have suggested that at least three tests for each domain would be important to provide adequate psychometric measurement (Kenny et al, 1988). The tests chosen for this study were those that were viewed as most relevant to assess neurocognitive response to treatment in a multisite trial. Our strategy was to include those cognitive functions that are most severely impaired in schizophrenia and are the most strongly associated with functional outcome. These selection criteria would be less likely to include tests of motor function and visual and auditory processing. Further, the practicality of assessing patients at 57 different sites eliminated the possibility of more time- and resource-intensive tests.

The second limitation of the 'single factor' conclusion is that the current results may describe the structure of neurocognitive *performance* in schizophrenia for the battery of tests used in this study, but owing to the absence of a control group, cannot be interpreted to explain the structure of the *deficits* in these patients. However, analyses of the cognitive *deficits* found in patients compared to controls suggest that impairment on a single common factor accounts for the vast majority of the between-group difference on the Wechsler Adult Intelligence Scale (Wechsler, 1997a) and Wechsler Memory Scale (Wechsler, 1997b) subtests, with more specific domains contributing very little unique between group variance (Dickinson et al, 2004). More recent approaches using CFA suggest a more generalized latent structure of cognitive ability in schizophrenia than in healthy subjects (Dickinson et al, in press).

The results of this previous study and our current findings suggest that a single dimension of cognitive performance is the preferred way to summarize cognitive impairment in schizophrenia using the test battery from this study. These data thus support the use of a single composite score derived from five domain scores as the primary neurocognitive outcome measure for the treatment analyses that will be performed on the CATIE data set. As the confirmatory factor analytic results suggest that a five-factor solution would be another appropriate approach to summarizing the CATIE neurocognitive variables, and since previous work on the effect of antipsychotic medications on cognition suggests

that some domains may demonstrate more improvement than others (Harvey and Keefe, 2001; Woodward et al, 2005), secondary analyses will focus on the five domain scores. However, a conclusion of variable treatment response across domains of cognitive deficit in patients with schizophrenia will need to be supported by evidence that these treatment responses are indeed 'differential' (Chapman and Chapman, 1973), in that there are significant differences among the cognitive changes on tasks that are matched for task complexity, a point that has long been emphasized in schizophrenia research (Chapman and Chapman, 1973).

It is certainly possible that a cotreatment for schizophrenia may target a single domain of cognitive function (eg working memory). However, the data from this study have implications for future studies assessing the potential procognitive effects of new compounds for schizophrenia. These trials will be challenged to balance an emphasis on improving individual domains of cognitive function, on the one hand, and general cognitive effects, on the other. As general cognitive factors may be more important in the determination of functional outcomes than individual domains (Green et al, 2000), the improvement of cognition summarized across multiple domains, as measured by a composite score in this study, may be particularly important for improving patient functioning.

As expected, the correlations of overall cognition as measured by the composite score with age, education, and duration of illness were about 0.3, defined by Cohen (1977) as 'medium' effect sizes. Patients who completed the entire test battery were slightly younger than those with greater than three missing tests. This difference was surprisingly small given that the age of patients in the study extended up to age 65. Older patients may be particularly challenged to complete a 90-min test battery. These relations with age are in line with previous studies of patients with schizophrenia and healthy controls (Heaton et al, 2001).

The magnitude of the correlations between cognition and clinical symptoms in schizophrenia has been inconsistent, but generally small (Goldberg et al, 1993; Addington et al, 1991; Hughes et al, 2003; Strauss, 1993; Mohamed et al, 1999; Bilder et al, 2000; Gladsjo et al, 2004). This finding, which may contradict clinicians' impressions, could potentially be explained by the possibility that patients who agree to participate in research studies do not reflect the general population of patients with schizophrenia, especially for those studies with extensive and demanding neurocognitive batteries. In the current 'all-comer' study, however, in which the vast majority of patients provided enough data to derive a meaningful composite score, the correlations of neurocognitive deficits with positive symptoms were near zero in all neurocognitive domains. In addition, the correlations between neurocognitive deficits and general symptoms were also low, ruling out the notion that neurocognitive deficits can be ascribed to symptoms such as mood and anxiety. As expected, however, the correlations with negative symptoms were in the small to medium range, consistent with previous studies (Addington et al, 1991; Buchanan et al, 1997; Hughes et al, 2003; Tamlyn et al, 1992; Bilder et al, 2000). While most evidence suggests that the relationship between negative symptoms and neurocognitive deficits is not causal (Harvey et al, 1996), the data from this study clearly indicate that in the general population of

patients with schizophrenia, neurocognitive deficits are certainly not caused by positive symptoms, as they are completely uncorrelated. It is possible that since the sample in this study is large and diverse, relationships that exist in more homogeneous samples have been hidden. Clustering solutions will be considered for future studies as the data from this project are further analyzed by the investigators. Overall, a more mature view of negative symptoms and neurocognitive deficits may emerge from considering both as observable signs of underlying pathology, and given their nonoverlap (ie 90% of variance not shared between the clinical symptom and neurocognitive assessment approaches), efforts to better understand their unique neural systems bases may be encouraged.

An important consideration for the assessment of neurocognition in clinical trials is the amount of testing that needs to be completed to provide sufficient neurocognitive information. The CATIE neurocognitive test battery requires about 75–90 min of testing time. However, briefer test batteries may reduce patient burden, tester time, and missing data points. Correlational analyses of the data in this study suggested that eliminating certain tests from the composite score did not significantly degrade the amount of information that the composite score provided. Even a short battery of tests—digit symbol, list learning, grooved pegboard, letter–number sequencing, verbal fluency, and a maze task (about 35 min of testing)—enabled the development of a summary score that accounted for 93% of the variance in the composite score developed from all tests, and several alternate approaches suggested that other indices derived from five or six tests selected to cover similar cognitive domains yield similarly accurate estimates. It should be noted that these estimates are a product of 'part-whole correlations' that are inflated by the perfect correlation between error variance in the summary score and error variance in each test measure. However, they support the notion that briefer test batteries are capable of providing researchers with almost as much information as longer batteries for calculating an overall composite score. As they reduce the time and resources required to conduct an assessment of neurocognition, they may be preferred over longer batteries for the calculation of summary scores in clinical trials (Wilk et al, 2005; Keefe et al, 2004; Velligan et al, 2004). However, brief approaches to the assessment of neurocognition may be more limited in their ability to provide information about specific neurocognitive effects of treatment, which will require more detailed assessment across multiple neurocognitive domains, as recommended in the conclusions of the MATRICS consensus process (Green et al, 2004b). Investigators with hypotheses about a specific cognitive domain, however, may find it practical to combine broad cognitive assessment using briefer tools, together with more detailed examination of the domain(s) of interest.

The amount of testing required to assess these multiple domains may decrease the number of patients who can provide complete data. In the CATIE schizophrenia trial, which utilized a design that encouraged a broad spectrum of patients to enroll, a large percentage of patients could be tested. About 3% of patients could not be tested at all. However, of those patients who were tested, 17% had one of the nine measures missing. More than 90% of patients provided data for at least six of the nine measures from the battery, suggesting that the overwhelming majority of

patients can complete the majority of neurocognitive tests. The composite scores that were derived from patients with only six of the nine measures had very high correlations with summary scores that included all tests. Thus, while meaningful neurocognitive summary scores can be collected on almost all patients entered into a trial such as this one, the ability to assess certain specific domains of neurocognitive performance is more variable.

A closer inspection of the pattern of missing data suggests that computerized tests may be more susceptible than noncomputerized tests to fail to provide meaningful results. The reasons for missing data were largely attributed to invalid scores and computer test malfunction rather than patient uncooperativeness, and a much greater percentage of patients did not complete one of the computerized tests compared to the noncomputerized tests. A composite score developed from all tests except the computerized tests included full data from 97% of the 1332 patients with fewer than four missing tests, and accounted for 95% of the variance in the composite score developed from all tests. Thus, in this battery of tests with computerized and noncomputerized tests, the reduction in sample size was greatest with the computerized tests, and these tests provide a relatively modest amount of unique information to the battery. More recent methods of computerized testing, including immediate data transfer and the availability of immediate feedback on test scores and technical difficulties (Barua et al, 2005) might diminish the impact of this problem. It is also possible that some of the failures may be attributable to patients finding the demands of selected computerized tests more challenging (eg the continuous performance test (CPT) involved a counter-intuitive 'finger lift' rather than 'button press' response, and this has been seen as a problem by experienced examiners). Furthermore, rather than an integrated package, the CATIE battery was comprised of computerized tests from different vendors, which may be more prone to missing data.

About 13% of the variance in the neurocognitive composite score was due to differences across sites. This result is consistent with previous clinical trials (Harvey et al, 2003, 2004), and has serious implications for trials assessing neurocognitive performance variables. The CATIE Project was designed to include a variety of different types of sites: academic sites; clinical research organizations, and community mental health sites with little clinical trials experience. There was also a wide disparity across sites in terms of geographical location and demographic characteristics. Thus, this design may have maximized between-site differences. These differences were not explained by differences in reading level, which estimates educational achievement. It is possible that the intersite differences may have been due to variable test administration and scoring. All testers were individually certified by CATIE personnel and were required to pass interim certification tests (Keefe et al, 2003); in addition, all sites were required to have at least one tester receive face-to-face certification. However, these procedures may not have been sufficient to guarantee consistent intersite test methods, emphasizing the need for intensive training and certification procedures for clinical trials involving cognitive assessment. The treatment analyses from this trial will thus attempt to control for intersite differences statistically to reduce potential error variance.

At the completion of testing at baseline, each patient was asked whether he or she had ever previously taken each of the tests. Only a small percentage of patients reported having taken the tests before, and experience had little impact on performance. These data challenge the notion that neurocognitive data in clinical trials are often compromised by variable exposures to the tests across patients. It is possible that a larger percentage of patients had taken the tests before, but they do not remember doing so. However, if they could not recall taking the tests, it is unlikely that they would have benefited substantially from their previous experience with the test.

In summary, data from 1332 patients assessed with neurocognitive tests in an 'all-comer' antipsychotic treatment trial suggest that neurocognitive deficits are not isolated to patients who enlist for research studies at academic institutions or highly filtered samples for clinical trials, nor do deficits appear more severe in the more general population of people with schizophrenia. Inspection of the factor structure of the neurocognitive measures supported the notion that these measures are moderately correlated with one another, and that a single factor model comprised of five domain scores best described the data. A single factor accounted for a large portion of the variance in performance that may be measured efficiently by a small number of tests. Although some patients did not complete all tests, the loss of data was not significant, as meaningful composite scores could be determined for more than 90% of patients who entered into this 'all-comer' trial. These neurocognitive deficits are unrelated to positive symptoms even in a sample of patients who were not extensively screened for suitability for a clinical trial. Subsequent publications from the CATIE Project will demonstrate the utility of the cognitive tests used here to predict function and improvement in function. Similarly, these data will be used to assess the importance of cognition in determining the relative effectiveness of treatment with typical and atypical drugs, thus providing a benchmark for future studies with the CATIE battery.

REFERENCES

Addington J, Addington D, Maticka-Tyndale E (1991). Cognitive functioning and positive and negative symptoms in schizophrenia. *Schizophr Res* 5: 123–134.

Arbuckle JL (2003). *Amos 5.0 Update to the Amos User's Guide*. SPSS: Chicago, IL.

Barua P, Bilder R, Small A, Sharma T (2005). Standardization and cross-validation study of cogtest an automated neurocognitive batter for use in clinical trials of schizophrenia. *Schizophr Bull* 31: 318.

Benton AL, Hamscher K (1978). *Multilingual Aphasia Examination Manual (revised)*. University of Iowa: Iowa City, IA.

Bilder RM, Goldman RS, Robinson D, Reiter G, Bell L, Bates JA et al (2000). Neuropsychology of first-episode schizophrenia: initial characterization and clinical correlates. *Am J Psychiatry* 157: 549–559.

Bowie CR, Serper MR, Riggio S, Harvey PD (2005). Neurocognition, symptomatology and functional skills in older alcohol-abusing schizophrenia patients. *Schizophr Bull* 31: 175–182.

Brandt J, Benedict R (1991). *Hopkins Verbal Learning Test*. Psychological Assessment Resources: Lutz, FL.

Browne M, Cudeck R (1993). Alternative ways of assessing model fit. In: Bollen K, Long J (eds). *Testing Structural Equation Models*. Sage Publications: Newbury Park, CA. pp 136–162.

Buchanan RW, Strauss ME, Breier A, Kirkpatrick B, Carpenter WT (1997). Attentional impairments in deficit and nondeficit forms of schizophrenia. *Am J Psychiatry* 154: 363–370.

Chapman LJ, Chapman JP (1973). *Disordered Thought in Schizophrenia*. Appleton-Century-Crofts: New York.

Cohen J (1977). *Statistical Power Analysis for the Behavioral Sciences (Revised)*. Academic Press: Orlando, FL.

Cornblatt BA, Keilp JG (1994). Impaired attention, genetics, and the pathophysiology of schizophrenia. *Schizophr Bull* 20: 31–46.

Dickinson D, Iannone VN, Wilk CM, Gold JM (2004). General and specific cognitive deficits in schizophrenia. *Biol Psychiatry* 55: 826–833.

Dickinson D, Ragland JD, Calkins ME, Gold JM, Gur RC (in press). A comparison of cognitive structure in schizophrenia patients and healthy controls using confirmatory factor analysis. *Schizophr Res*.

First MB, Spitzer RL, Gibbon M, Williams JBW (1996). *Structured Clinical Interview for Axes I and II DSM-IV Disorders-Patient Edition (SCID-I/P)*. Biometrics Research Department, New York State Psychiatric Institute: New York.

Gladsjo JA, McAdams LA, Palmer BW, Moore DJ, Jeste DV, Heaton RK (2004). A six-factor model of cognition in schizophrenia and related psychotic disorders: Relationships with clinical symptoms and functional capacity. *Schizophr Bull* 30: 739–754.

Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR (1997). Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. *Arch Gen Psychiatry* 54: 159–165.

Goldberg TE, Gold JM, Greenberg R, Griffin S, Schulz SC, Pickar D et al (1993). Contrasts between patients with affective disorders and patients with schizophrenia on a neuropsychological test battery. *Am J Psychiatry* 150: 1355–1362.

Green AI, Tohen MF, Hamer RM, Strakowski SM, Lieberman JA, Glick I et al (2004a). First episode schizophrenia-related psychosis and substance use disorders: acute response to olanzapine and haloperidol. *Schizophr Res* 66: 125–135.

Green MF, Kern RS, Braff DL, Mintz J (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the 'right stuff'? *Schizophr Bull* 26: 119–136.

Green MF, Nuechterlein KH, Gold JM, Barch D, Cohen J, Essock S et al (2004b). Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. *Biol Psychiatry* 56: 301–307.

Green MF (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry* 153: 321–330.

Guy W (1976). *Early Clinical Drug Evaluation (ECDEU), Assessment Manual*. National Institute of Mental Health: Rockville, MD.

Harvey PD, Green MF, McGurk SR, Meltzer HY (2003). Changes in cognitive functioning with risperidone and olanzapine treatment: A large-scale, double blind, randomized study. *Psychopharmacology* **169**: 404–411.

Harvey PD, Keefe RSE (1997). Cognitive impairment in schizophrenia and implications of atypical neuroleptic treatment. *CNS Spectr* **2**: 1–11.

Harvey PD, Keefe RSE (2001). Studies of cognitive change in patients with schizophrenia following novel antipsychotic treatment. *Am J Psych* **158**: 176–184.

Harvey PD, Lombardi J, Leibman M, White L, Parrella M, Powchik P et al (1996). Cognitive impairment and negative symptoms in schizophrenia: a prospective study of their relationship. *Schizophr Res* **22**: 223–231.

Harvey PD, Siu CO, Romano S (2004). Randomized, controlled, double-blind, multicenter comparison of the cognitive effects of ziprasidone versus olanzapine in acutely ill inpatients with schizophrenia and schizoaffective disorder. *Psychopharmacology* **172**: 324–332.

Heaton RK, Gladsjo JA, Palmer BW, Kuck J, Marcotte TD, Jeste DV (2001). Stability and course of neuropsychological deficits in schizophrenia. *Arch Gen Psychiatry* **58**: 24–32.

Heinrichs RW, Zakzanis KK (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* **12**: 426–444.

Hughes C, Kumari V, Soni W, Das M, Binneman B, Drozd S et al (2003). Longitudinal study of symptoms and cognitive function in chronic schizophrenia. *Schizophr Res* **59**: 137–146.

Hyman SE, Fenton WS (2003). What are the right targets for psychopharmacology? *Science* **299**: 350–351.

Kay SR, Fiszbein A, Opler LA (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* **13**: 261–276.

Keefe RSE, Eesley CE, Poe MP (2005). Defining a cognitive function decrement in schizophrenia. *Biol Psychiatry* **57**: 688–691.

Keefe RSE, Goldberg TE, Harvey PD, Gold JM, Poe M, Coughenour L (2004). The brief assessment of cognition in schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr Res* **68**: 283–297.

Keefe RSE, Mohs RC, Bilder RM, Harvey PD, Green MF, Meltzer HY et al (2003). Neurocognitive Assessment in the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project schizophrenia trial: development, methodology and rationale. *Schizophr Bull* **29**: 45–55.

Kenny DA, Kashy DA, Bolger N (1988). Data analysis in social psychology. In: Gilbert D, Fiske S, Lindzey G (eds). *Handbook of Social Psychology*, 4th edn. McGraw-Hill: Boston. pp 233–265.

Kerr SL, Neale JM (1993). Emotion perception in schizophrenia: specific deficit or further evidence of generalized poor performance? *J Abnorm Psychol* **102**: 312–318.

Kongs SK, Thompson LL, Iverson GL, Heaton RK (2000). *Wisconsin Card Sorting Test-64 Card Version Professional Manual*. Psychological Assessment Resources, Inc.: Lutz, FL.

Lafayette Instrument Company (1989). *Grooved Pegboard Instruction Manual, Model 32025*. Lafayette Instrument Company: Lafayette, Indiana.

Lyons-Warren A, Lillie R, Hershey T (2004). Short and long-term spatial delayed response performance across the lifespan. *Dev Neuropsych* **26**: 661–678.

Mohamed S, Paulsen JS, O'Leary D, Arndt S, Andreasen N (1999). Generalized cognitive deficits in schizophrenia: a study of first-episode patients. *Arch Gen Psychiatry* **56**: 749–754.

Nuechterlein KH, Barch DM, Gold JM, Goldberg TE, Green MF, Heaton RK (2004). Identification of separable cognitive factors in schizophrenia. *Schizophr Res* **72**: 29–39.

Saykin AJ, Gur RC, Gur RE, Mozley PD, Mozley LH, Resnick SM et al (1991). Neuropsychological function in schizophrenia. Selective impairment in memory and learning. *Arch Gen Psychiatry* **48**: 618–624.

Slade E, Salkever D, Rosenheck R, Swanson J, Swartz M, Shern D et al (2005). Cost-sharing requirements and access to mental health care among medicare enrollees with schizophrenia. *Psychiatr Serv* **56**: 960–966.

Stirling J, Lewis S, Hopkins R, White C (2005). Cannabis use prior to first onset psychosis predicts spared neurocognition at 10-year follow-up. *Schizophr Res* **75**: 135–138.

Strauss ME (1993). Relations of symptoms to cognitive deficits in schizophrenia. *Schizophr Bull* **19**: 215–231.

Stroup TS, McEvoy JP, Swartz MS, Byerly MJ, Glick ID, Canive JM et al (2003). The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Project: schizophrenia trial design and protocol development. *Schizophr Bull* **29**: 15–31.

Swanson JW, Swartz MS, Van Dorn RA, Elbogen EB, Wagner HR, Rosenheck RA et al (in press). A national study of violent behavior in persons with schizophrenia. *Arch Gen Psychiatry*.

Tamlyn D, McKenna PJ, Mortimer AM, Lund CE, Hammond S, Baddeley AD (1992). Memory impairment in schizophrenia: its extent, affiliations and neuropsychological character. *Psychol Med* **22**: 101–115.

Tulsky DS, Price LR (2003). The joint WAIS-III and WMS-III factor structure: development and cross-validation of a six-factor model of cognitive functioning. *Psychol Assess* **15**: 149–162.

Velligan DI, DiCocco M, Bow-Thomas CC, Cadle C, Glahn DC, Miller AL et al (2004). A brief cognitive assessment for use with schizophrenia patients in community clinics. *Schizophr Res* **71**: 273–283.

Wechsler D (1974). *Wechsler Adult Intelligence Scale*, Revised edn. Psychological Corporation: San Antonio, TX.

Wechsler D (1991). *Wechsler Intelligence Scale for Children*, 3rd edn. Harcourt Publishers, NY.

Wechsler D (1997a). *Wechsler Adult Intelligence Scale*, 3rd edn. Psychological Corporation: San Antonio, TX.

Wechsler D (1997b). *Wechsler Memory Scale*, 3rd edn. Psychological Corporation: San Antonio, TX.

Wilk CM, Gold JM, Humber K, Dickerson F, Fenton WS, Buchanan RW (2004). Brief cognitive assessment in schizophrenia: Normative data for the Repeatable Battery for the Assessment of Neuropsychological Status. *Schizophr Res* **70**: 175–186.

Wilk CM, Gold JM, McMahon RP, Humber K, Lannone VN, Buchanan RW (2005). No, it is not possible to be schizophrenic yet neuropsychologically normal. *Neuropsychology* **19**: 778–786.

Woodward ND, Purdon SE, Meltzer HY, Zald DH (2005). A meta-analysis of neuropsychological change to clozapine, olanzapine, quetiapine, and risperidone in schizophrenia. *Int J Neuropsychopharm* **8**: 1–16.

APPENDIX I

CATIE Study Investigators Group includes: Lawrence Adler, MD, Clinical Insights, Glen Burnie, MD; Mohammed Bari, MD, Synergy Clinical Research, Chula Vista, CA; Irving Belz, MD, Tri-County/MHMR, Conroe, TX; Raymond Bland, MD, Southern Illinois University School of Medicine, Springfield, IL; Thomas Blocher, MD, MHMRA of Harris County, Houston, TX; Brent Bolyard, MD, Cox North Hospital, Springfield, MO; Alan Buffenstein, MD, The Queen's Medical Center, Honolulu, HI; John Burruss, MD, Baylor College of Medicine, Houston, TX; Matthew Byerly, MD, University of Texas Southwestern Medical Center at Dallas, Dallas, TX; Jose Canive, MD, Albuquerque VA Medical Center, Albuquerque, NM; Stanley Caroff, MD,

Behavioral Health Service, Philadelphia, PA; Charles Casat, MD, Behavioral Health Center, Charlotte, NC; Eugenio Chavez-Rice, MD, El Paso Community MHMR Center, El Paso, TX; John Csernansky, MD, Washington University School of Medicine, St Louis, MO; Pedro Delgado, MD, University Hospitals of Cleveland, Cleveland, OH; Richard Douyon, MD, VA Medical Center, Miami, FL; Cyril D'Souza, MD, Connecticut Mental Health Center, New Haven, CT; Ira Glick, MD, Stanford University School of Medicine, Stanford, CA; Donald Goff, MD, Massachusetts General Hospital, Boston, MA; Silvia Gratz, MD, Eastern Pennsylvania Psychiatric Institute, Philadelphia, PA; George T Grossberg, MD, St Louis University School of Medicine-Wohl Institute, St Louis, MO; Mahlon Hale, MD, New Britain General Hospital, New Britain, CT; Mark Hamner, MD, Medical University of South Carolina and Veterans Affairs Medical Center, Charleston, SC; Richard Jaffe, MD, Belmont Center for Comprehensive Treatment, Philadelphia, PA; Dilip Jeste, MD, University of California-San Diego, VA Medical Center, San Diego, CA; Anita Kablinger, MD, Louisiana State University Health Sciences Center, Shreveport, LA; Ahsan Khan, MD, Psychiatric Research Institute, Wichita, KS; Steven Lamberti, MD, University of Rochester Medical Center, Rochester, NY; Michael T Levy, MD, PC, Staten Island University Hospital, Staten Island, NY; Jeffrey Lieberman, MD, University of North Carolina School of Medicine, Chapel Hill, NC; Gerald Maguire, MD, University of California Irvine, Orange, CA; Theo Manschreck, MD, Corrigan Mental Health Center, Fall River, MA; Joseph McEvoy, MD, Duke University Medical Center, Durham, NC; Mark McGee, MD, Appalachian Psychiatric Healthcare System, Athens, OH; Herbert Meltzer, MD, Vanderbilt University Medical Center, Nashville, TN; Alexander Miller, MD, University of Texas Health Science Center at San Antonio, San Antonio, TX; Del D Miller, MD, University of Iowa, Iowa City, IA; Henry Nasrallah, MD, University of Cincinnati Medical Center, Cincinnati, OH; Charles Nemeroff, MD, PhD, Emory University School of Medicine, Atlanta, GA; Stephen Olson, MD, University of Minnesota Medical School, Minneapolis, MN; Gregory F Oxenkrug, MD, St Elizabeth's Medical Center, Boston, MA; Jayendra Patel, MD, University of Mass Health Care, Worcester, MA; Frederick Reimher, MD, University of Utah Medical Center, Salt Lake City, UT; Silvana Riggio, MD, Mount Sinai Medical Center-Bronx VA Medical Center, Bronx, NY; Samuel Risch, MD, University of California-San Francisco, San Francisco, CA; Bruce Saltz, MD, Henderson Mental Health Center, Boca Raton, FL; Thomas Simpatico, MD, Northwestern University, Chicago, IL; George Simpson, MD, University of Southern California Medical Center, Los Angeles, CA; Michael Smith, MD, Harbor—UCLA Medical Center, Torrance, CA; Roger Sommi, PharmD, University of Missouri, Kansas City, MO; Richard M Steinbook, MD, University of Miami School of Medicine, Miami, FL; Michael Stevens, MD, Valley Mental Health, Salt Lake City, UT; Andre Tapp, MD, VA Puget Sound Health Care System, Tacoma, WA; Rafael Torres, MD, University of Mississippi, Jackson, MS; Peter Weiden, MD, SUNY Downstate Medical Center, Brooklyn, NY; and James Wolberg, MD, Mount Sinai Medical Center, New York, NY.

APPENDIX II

List of scores generating queries or changed to 'missing'.

Test	Scores changed to 'missing'	Scores generating a query to the tester
WRAT III reading test	● Less than 37 if English was not the primary language: all cognitive data changed to missing	● Score less than 26
Controlled oral word association test		● Score greater than 30 for one letter
Hopkins verbal learning test		● Score of 0 on one of the three trials
WAIS-R digit symbol		● Score greater than 84
Grooved pegboard		● Score less than 5
Continuous performance test	● D-prime scores less than −1 ● Reaction time scores equal to 0 ● Finger ups for the number of 'Hits' greater than or equal to 10	
Visuospatial working memory	● Mean error for 'No Delay' greater than or equal to 20	

Note: A score of 0 for any test generated a query to the tester.