Temporal Stability and Moderating Effects of Age and Sex on CNTRaCS Task Performance

Milton E. Strauss[*,1], **Christopher J. McLouth**[1], **Deanna M. Barch**[2], **Cameron S. Carter**[3], **James M. Gold**[4], **Steven J. Luck**[5], **Angus W. MacDonald III**[6], **J. Daniel Ragland**[3], **Charan Ranganath**[5], **Brian P. Keane**[7,8], and **Steven M. Silverstein**[7]

[1]Department of Psychology, University of New Mexico, Albuquerque, NM; [2]Department of Psychology, Washington University, St Louis, MO; [3]Department of Psychiatry, University of California, Davis, CA; [4]Department of Psychiatry, Maryland Psychiatric Research Center, Baltimore, MD; [5]Department of Psychology, University of California, Davis, CA; [6]Department of Psychology, University of Minnesota, Minneapolis, MN; [7]Department of Psychiatry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, NJ; [8]Center for Cognitive Science, Rutgers University, New Brunswick, NJ

*To whom correspondence should be addressed; Department of Psychology, University of New Mexico, PO Box 3837, Corrales, Albuquerque, NM 87048, US; tel: 314-935-8729, fax: 314-935-8790, e-mail: milton.strauss@gmail.com

Research in schizophrenia has increasingly focused on incorporating measures from cognitive neuroscience, but little is known about their psychometric characteristics. Here, we extend prior research by reporting on temporal stability, as well as age and sex effects, for cognitive neuroscience paradigms optimized as part of the Cognitive Neuroscience Test Reliability and Clinical applications for Schizophrenia consortium. Ninety-nine outpatients with schizophrenia and 131 healthy controls performed 5 tasks assessing 4 constructs at 3 sessions. The constructs were (1) Goal maintenance (Dot Probe Expectancy [DPX] and AX continuous performance tasks [AX-CPT]); (2) Episodic memory (Relational and Item-Specific Encoding and Retrieval task [RiSE]); (3) Visual integration (Jittered Orientation Visual Integration task [JOVI]); and (4) Perceptual gain control (Contrast-Contrast Effect Task [CCE]). Patients performed worse than controls on all but the CCE, and the magnitude of these group differences was stable across sessions, with no sex differences observed. Improvements over sessions were seen for the AX-CPT, the DPX, and the JOVI though practice effects for the AX-CPT and the DPX were primarily present in older participants. For the AX-CPT and the JOVI, practice effects were larger for T1 to T2 than for T2 to T3. Age was associated with poor associative recognition on the RiSE and accuracy on the JOVI. Test-rest reliability ranged from poor for the JOVI threshold score to adequate to good for the DPX, AX-CPT, and JOVI accuracy measures, with RiSE and CCE measures in the moderate range. These results suggest that group differences in DPX, AX-CPT, RiSE, and JOVI are robust and consistent across repeated testing.

Key words: psychometrics/schizophrenia/cognitive neuroscience/goal maintenance/episodic memory/visual integration

Introduction

Experimental paradigms developed within cognitive neuroscience have become increasingly important in the field of clinical cognitive neuroscience. Such tasks have potential for measuring discrete, neurally dissociable impairments that may be suitable targets for intervention and/or which may serve as endophenotypic markers of risk for illness.[1,2] The study reported here is based on data from a series of multisite investigations of constructs and measures derived from cognitive neuroscience relevant to schizophrenia. The studies are part of the Cognitive Neuroscience Test Reliability and Clinical applications for Schizophrenia (CNTRaCS) consortium.[3] This article reports on the second study of these measures in a large new sample of patients with schizophrenia and a sample of psychiatrically healthy comparison participants tested 3 times. The study was concerned with the (1) temporal stability of differences between patients and controls (ie, differential practice effects); (2) associations of performance with age; (3) sex differences in task performance; and (4) practicality and tolerability. In addition to the stability of group differences in the absence of intervention, we examined the test-retest reliability of individual differences over the 3 testing occasions.

Stability of Performance in Schizophrenia Patient Groups

Temporal reliability of task performance among clinically stable patients is important to establish prior to intervention studies because practice effects may reduce intervention effects.[4–6] The effect of repeated assessment in clinically stable patients has primarily been examined using measures based in clinical neuropsychology. Stability on such measures varies across tasks[7,8] and across studies, with both lower stability[7,9] and comparable stability[10] reported when comparing patients with controls. However, this is the first study to our knowledge to examine different aspects of stability (both stability of group differences and individual differences) on tasks derived from cognitive neuroscience across several assessments in persons with schizophrenia and controls. Importantly, there is some evidence[5,11] that use of multiple baselines may ameliorate practice effects resulting from learning of task structure and response demands, and this approach was implemented in this study. The idea of multiple baselines is that participants' first experience with a task results in the greatest learning, and therefore the greatest practice effects on the next testing, whereas change from a second baseline to any subsequent point as a function of practice is likely to be reduced.

Aging and Schizophrenia

A number of investigators have noted similarities between the psychological deficits typically seen in schizophrenia and the performance of older healthy individuals[12–15] for an early review. Age-associated impairments for schizophrenia appear to be more generalized among those persistently hospitalized, whereas in outpatients age-related impairments are most prominent for memory, attention, and processing speed measures.[16–18] This suggests that variation in age may need to be accounted for in the evaluation of cognitively focused interventions or studies of cognition in schizophrenia, and that age may influence the extent to which practice effects are observed. Thus, we investigated main effects of age and interactions between age and practice effects.

Sex Differences in Schizophrenia

Sex differences in schizophrenia have been reported for clinical characteristics, prognosis, and neuroimaging, but only sometimes for cognitive performance. When sex differences are found, deficits are greater for men in many studies.[19–23] However, most studies of cognition in schizophrenia have been underpowered for detecting sex differences. Samples are predominantly male and the power of a study is substantially affected by the size of the smaller group.[19,20] There are both public health[20] and clinical research reasons for adequately sampling and comparing both sexes, including suggestions of sex differences in treatment response.[24,25] This study was designed so that approximately half the participants in each group were women, allowing sufficient power to examine sex effects on cognition in schizophrenia.

The paradigms studied here were selected through a consensus-based process (CNTRICS) with the requirement that they measured discrete cognitive processes with evidence from basic cognitive neuroscience linking performance to distinct cognitive and neural systems (see Henderson et al,[26] Barch et al,[27] Ragland et al[28], and Silverstein et al[29] for reviews of the validity of these tasks as measures of specific constructs) rather than a common factor or generalized deficit. The 4 tasks and associated constructs were (1) Dot Probe Expectancy (DPX), a nonverbal variant of the AX continuous performance task (AX-CPT)[30] to assess *goal maintenance in working memory*[26]; (2) Relational and Item-Specific Encoding and Retrieval (RiSE), a measure of distinct episodic memory *encoding and retrieval* processes[28]; (3) Jittered Orientation Visual Integration (JOVI), an assay of *visual integration* or perceptual organization[29]; and (4) Contrast-Contrast Effect Task (CCE), an index of *perceptual gain control* or center-surround suppression.[27] A second version of a goal maintenance task, the AX-CPT, which uses letter stimuli, was included to provide a source of convergent validation for the DPX and to follow up on previous studies.[31] Analyses were designed to determine whether these tasks were appropriate for the assessment of change in these specific cognitive processes in clinical trials.

Method

Participants

Outpatients and partially hospitalized individuals with schizophrenia and psychiatrically healthy adults (18–65 years of age) were recruited by the CNTRaCS consortium from University of California—Davis, Maryland Psychiatric Research Center at the University of Maryland, University of Medicine and Dentistry of New Jersey, University of Minnesota—Twin Cities, and Washington University in St Louis (see online supplementary materials for site Ns). The total number of patients tested was 103 in the schizophrenia group and 132 for the healthy controls. The study was approved by the Institutional Review Board of each participating institution; written informed consent was obtained from each participant.

Inclusion and exclusion criteria were (1) no history of significant head trauma or neurological disease; (2) no history of mental retardation or pervasive developmental disorder; (3) and no history of substance dependence in the last 6 months or substance abuse in the last month. All participants were English speakers and scored ≥6 on the Wechsler Test of Adult Reading (WTAR).[32] Each participant was also required to pass alcohol and drug testing on every visit to the laboratory.[26] Patients were excluded if their medication had changed over the

preceding month. For controls, additional criteria included no history of any psychotic or bipolar disorder, no current major depression, and no current psychotropic or cognitive-enhancing medications. There were no exclusion criteria based on family history. A master-level clinician conducted diagnostic assessments using the Structured Clinical Interview for DSM-IV-TR[33] and the 24-item Brief Psychiatric Rating Scale.[34,35] Details regarding interviewer training and establishing and monitoring reliability are available in the online supplementary materials.

Procedure

As noted above, 5 cognitive tasks intended to measure 4 different constructs were administered to each participant. Each of these tasks has been described in detail in earlier reports, and additional details are provided in table 1 and the online supplementary materials, including descriptions of task length. The software for these tasks and testing procedures also are available for use by other investigators through the CNTRaCS Web site http://cntracs.ucdavis.edu/task. The order of task administration within a session for each participant was the same as the order in which the tasks are listed, with the exception of the goal maintenance tasks, which were counterbalanced across participants. We used a fixed task order in order to facilitate individual differences analyses of the relationships between task performance and other measures (eg, function, symptoms), which will be the focus of

other publications. All tasks were administered 3 times. The time between the first (T1) and second (T2) assessments was planned to be 7 days ($M = 7.24$ days, range 4–14) and between the second and third (T3) 14 days ($M = 14.39$ days, range 10–29). Similar to the MATRICS,[36] we also collected tolerability ratings from participants for the RiSE, JOVI, and CCE (ratings are not available for the DPX and the AX-CPT due to a technical problem with the part of the Eprime script that was to present the rating questions). The ratings were made using a 7-point visual analog scale, with 1 anchored with a frowning face and the word "Unpleasant" and 7 anchored with a smiling face and the word "Pleasant."

Data Analysis Overview

Multilevel modeling (MLM) was used to analyze performance because it permits the inclusion of data from participants with at least 2 data points and accounts for the actual assessment intervals for each subject. Models were built independently for each cognitive task using established procedures.[29] Assessment occasion (time) was the level-1 predictor, which was centered at the first testing occasion. The between person (level-2) predictors were diagnosis, age (grand-mean centered), sex, and all higher order interactions. A missing data analysis indicated that diagnostic group, age, sex, parental and own SES, and site did not predict missing values and so MLM analyses were conducted assuming random missing data.

Table 1. Description of CNTRaCS Tasks

Construct	Task Name	Number of Trials	Dependent Measure(s)
Goal maintenance: The processes involved in activating task related goals or rules based on endogenous or exogenous cues, actively representing them in a highly accessible form, and maintaining this information over an interval during which that information is needed to bias and constrain attention and response selection	Dot Probe Expectancy (DPX)[26] AX-CPT[30,36]	144 (104 "A-X", 16 "B-X", 16 "A-Y", 8 "B-Y") 144 (104 "A-X", 16 "B-X", 16 "A-Y", 8 "B-Y")	d'-context (AX hits vs BX false alarms) d'-context (AX hits vs BX false alarms)
Episodic memory encoding and retrieval: The processes involved in memory for stimuli/ elements and how they were associated with coincident context, stimuli or events	Relational and Item-Specific Encoding (RiSE)[28]	36 Item Encoding and 36 Relational Encoding; 144 Item Recognition (½ new, ½ old), 18 Associative Recognition	Corrected Recognition (hits minus false alarms) for (a) item recognition: (1) item encoding and (2) the relational encoding condition and (b) associative recognition
Visual integration: The processes linking the output of neurons – that individually code local (typically, small) attributes of a scene - into global (typically, larger) complex structure, more suitable for the guidance of behavior	Jittered Orientation Visual Integration (JOVI)[29]	288 (48 at each of 6 "jitter" levels)	Threshold defined using a sigmoidal function (online supplementary materials); percent accuracy
Gain control: The processes whereby neurons adapt their response levels to take into account their immediate context, in order to make best use of a limited dynamic signaling range	Contrast-Contrast Effect (CCE)[27]	180 in no surround, 180 in surround, each of which included 20 catch trials	Difference between performance in the surround and no surround condition

The first step in each analysis was to evaluate a fully unconditional model against which more constrained models were compared. Time was then added as a fixed effect and compared with a model that allowed the effect of time to vary randomly using the likelihood ratio test. In each case, permitting the slope of time to be an individual difference variable improved model fit. Next, a model including all other main effects and interactions was estimated. The diagnosis by age and diagnosis by time interactions were retained in all models. Other interactions were removed if nonsignificant before estimating the final model. The final models for each task had the same main effects but could differ in regard to the interactions. The main effects of group and age were tested at T1. Significant interactions with Time indicated that main effects changed over testings. Alpha was set at $P < .05$ for effects of diagnosis because these tests were replications of previous findings.[3] For other factors in the MLM analyses, alpha for the nominal .05-level was $P < .00625$ (.05/8) because there were a total of 8 dependent variables. Repeated Measures ANOVA was used to test expected group by task condition interactions, with alpha set at $P < .05$. All of these analyses were performed using SPSS Version 19. For analyses with participants with data from all 3 assessment points (Ns = 207–217), power exceed .9 for a medium effect size ($r = .3$)[37] for all comparisons.[38,39]

The skewness and kurtosis values for all dependent variables are shown in online supplementary table S1. With a relatively large sample size such as ours, even relatively small deviations from normality are significant. As such, prior work has suggested that variables with skewness values greater than ±2 should be transformed. None of the variables had skewness values above 2. However, it is standard in the field to log transform threshold scores and thus we did so for the JOVI threshold scores.

Results

Participants' Characteristics

Five participants (4 patients and 1 control) were dropped from the analyses. One patient and 1 control were dropped because their schedules did not allow them to be tested at the appropriate time intervals and 2 patients were dropped because they did not meet inclusion/exclusion criteria. Characteristics of the final sample of patients and controls are summarized in table 2. The groups were recruited to be matched for age, sex, and race/ethnicity. Patients, as expected, had lower personal socioeconomic status (SES), but parental SES did not differ between groups. Comparisons of demographic difference across sites indicated significant between-site variability for age ($F(4,220) = 30.25$, $P < .001$) and for both SES measures (Own: $F(4, 220) = 3.71$, $P = .006$; Parental: $F(4, 213) = 8.30$, $P < .001$).

Descriptive statistics for all task performance variables are presented in table 3. In addition, online supplementary table S2 shows the maximum and minimum scores for all variables, as well as the percentages of individuals at floor or ceiling at each testing point using the same criteria as for MATRICS.[36] A nested-model deviance test indicated that a 3-level model using site did not provide a better fit for the data than a 2-level model that ignores research site for any of the dependent variables (largest $X^2 = 0.714$, df = 1, $P = .398$). This suggested that performance was not different across research site, so site was not included in further analyses. Results of the statistical analyses are organized by group differences, changes over repeated testings, and association with aging. Interactions are considered within the context of group and testing effects. Only statistically significant effects are described. For complete results of all MLM and RM-ANOVA analyses, please see online supplementary materials.

Group Differences

Patients performed more poorly than normal controls on all but one task (table 3 and online supplementary tables S3–S6). Patient-control differences were comparable for both indices of *Goal Maintenance in Working Memory* (DPX: $b = −0.772$, $t(220) = −6.43$, $P < .001$; AX-CPT: $b = −0.660$, $t(220) = −6.11$, $P < .001$; the correlations between DPX and AX-CPT were between .63 and .80 across groups and testing sessions). There were 3

Table 2. Demographic Data for Schizophrenia and Control Groups

Variable	Patient, $n = 99$ Mean (SD)	Control, $n = 131$ Mean (SD)	Group Comparison
Age (years)	40.4 (11.7)	38.4 (12.3)	$t = −1.24$, $P = .215$
Gender (% males)	56.6	52.7	$X^2 = 0.35$, $P = .557$
Race (% Euro-American)	60.0	48.1	$X^2 = 3.00$, $P = .083$
Personal SES	24.0 (9.0)	33.1 (11.5)	$t = 6.73$, $P < .001$
Parental SES	45.6 (12.8)	43.5 (12.5)	$t = −1.21$, $P = .227$

Note: SES, socioeconomic status. SES is measured by the Barratt Simplified Measure of Social Status based on the Hollingshead Index.

Table 3. Means and SDs of Performance Measures for Patients With Schizophrenia and Healthy Controls

Task	Time 1[a] Patient n	Patient M (SD)	Control n	Control M (SD)	Time 2 Patient n	Patient M (SD)	Control n	Control M (SD)	Time 3 Patient n	Patient M (SD)	Control n	Control M (SD)
DPX	92	2.30 (1.25)	127	3.21 (0.88)	90	2.58 (1.10)	122	3.30 (0.83)	86	2.73 (1.15)	119	3.35 (0.80)
AX-CPT	95	2.58 (1.02)	128	3.34 (0.87)	93	2.90 (1.07)	123	3.48 (0.73)	88	2.96 (1.07)	122	3.50 (0.78)
CCE	98		128		95		123		94		123	
Difference		6.63 (11.43)		8.10 (10.64)		6.63 (12.55)		8.33 (8.85)		8.13 (9.09)		8.65 (9.15)
JOVI	92		130		92		128		94		122	
Threshold[b]		1.12 (0.10)		1.09 (0.07)		1.10 (0.10)		1.07 (0.08)		1.10 (0.10)		1.06 (0.09)
Proportion correct		0.69 (0.08)		0.72 (0.08)		0.72 (0.09)		0.72 (0.08)		0.73 (0.09)		0.77 (0.09)
RiSE	96		127		95		123		92		123	
AR		0.33 (0.24)		0.56 (0.20)		0.32 (0.27)		0.56 (0.25)		0.26 (0.28)		0.52 (0.23)
IRIE		0.65 (0.26)		0.84 (0.11)		0.66 (0.23)		0.85 (0.11)		0.70 (0.19)		0.84 (0.11)
IRAE		0.61 (0.27)		0.83 (0.11)		0.62 (0.22)		0.83 (0.14)		0.65 (0.21)		0.82 (0.13)

Note: DPX, Dot Probe Expectancy; AX-CPT, AX continuous performance test; RiSE, Relational and Item-Specific Encoding Task; AR, Associative Recognition; IRIE, Item Recognition, Item Encoding; IRAE, Item Recognition, Relational Encoding; CCE, Contrast-Contrast Effect; JOVI, Jittered Orientation Visual Integration.

[a] The time between first and second testing averaged 7.2 days (range 4–14) and between second and third testing averaged 14.4 days (range 10–29).

[b] Threshold values were transformed to conform to normality (see text). Larger values correspond to worse performance.

dependent variables for the *RiSE*: (1) item recognition for item encoding; (2) item recognition for relational encoding; and (3) and associative recognition for relational encoding. All 3 were impaired in patients relative to controls (item encoding: $b = −0.168$, $t(213) = −8.43$, $P < .001$; relational encoding, $b = −0.189$, $t(217) = −9.25$, $P < .001$; associative recognition, $b = −0.226$, $t(217) = −9.16$, $P < .001$). As anticipated,[29] patients performed most poorly for item recognition in the relational encoding vs item encoding condition ($P < .02$). As shown in table 1, the *JOVI* was indexed by both log transformed *JOVI* threshold scores and overall accuracy (proportion correct). At each time point, the proportion correct improved for both groups as the contour elements became less jittered ($P < .001$) indicating that participants in both groups were sensitive to this manipulation of perceptual organization. More importantly, patients performed more poorly on both indicators suggesting people with schizophrenia have more difficulty integrating spatially separated visual elements into a single, coherent shape (for threshold $t(213) = 2.52$, $P = .03$; for accuracy, $b = −0.034$, $t(221) = −3.46$, $P = .001$). Participants on the whole were subject to the CCE illusion, but the 2 groups did not differ, as indexed by the difference in the contrast ratings between the no surround and surround conditions ($b = −0.90$, $t(217) = −0.75$, $P = .453$).

Magnitude of Group Difference Effects. Using Cohen's[40] convention, the ESs for the DPX and AX-CPT were medium (Cohen's $d ≥ .62$), while those for all 3 RiSE measures were large ($d ≥ 1.16$). The ES for JOVI threshold difference was small ($d = .38$), while that for the accuracy measure was medium ($d = .54$). The CCE effect size was small ($d = .13$). With the exception of the JOVI threshold and CCE difference scores, diagnosis is associated with more variance than the time factor (practice effect).

Repeated Testing

Improvements in performance over time were found for the AX-CPT ($b = 0.057$, $t(186) = 3.39$, $p = .001$), the DPX (DPX: $b = 0.073$, $t(190) = 4.07$, $P < .001$). However, this pattern was seen principally among older persons in both groups: regressing age on the difference between T3 and T1 $d′$ scores indicated that improvements were linearly associated with age (standardized b DPX = *.27*; AZ-CPT = .25, $P < .002$; figure 1). The only other task for which there was a practice effect was the JOVI. Both groups improved over testings (for accuracy, $b = 0.011$, $t(194) = 8.61$, $P < .001$, for threshold, $b = −0.008$, $t(253) = −3.98$) and there was no interaction between testing and diagnostic group, indicating that the magnitude of group differences was stable across sessions.

To clarify when practice effects were occurring and to evaluate the potential efficacy of the repeated baseline approach, we compared the gains between T1 and
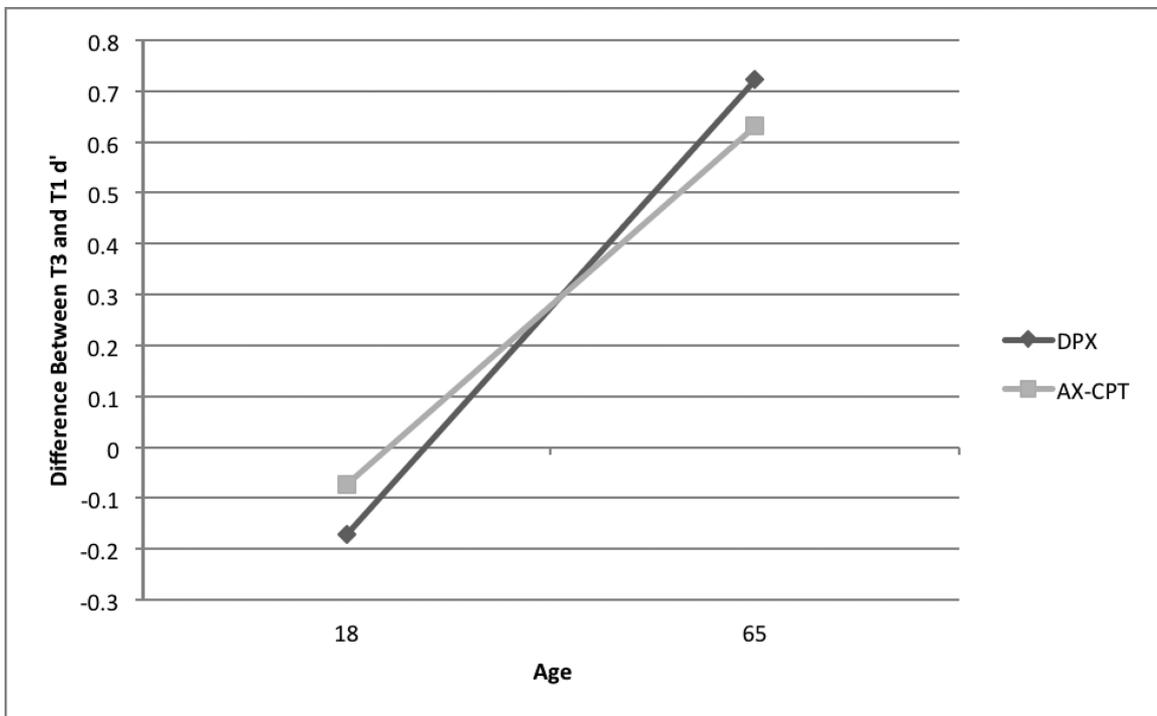
Fig. 1. Relationship of age to improvement in performance between the first and third testing occasions (patients and controls combined).

T2 with those between T2 and T3 for the d' and proportion correct measures. There was no differential improvement between the pairs of time points for DPX d', but the improvement for AX-CPT d' was significantly larger between the first 2 sessions than the second and third (M difference = 0.21 vs 0.04, $P < .025$). There was a comparable differential practice effect for JOVI proportion correct, with a greater practice effect improvement between the first 2 sessions (M difference = .03 vs .01, $P < .038$), that is, practice effects AX-CPT and JOVI were greater for between T1 and T2 than between T2 and T3 though the effect sizes are small (AX-CPT $r = .16$; JOVI accuracy $r = .19$).

Sex

Sex did not account for significant variance on any cognitive task (largest $t = -0.96$, $P = .338$). Further, there were no significant interactions between age and diagnosis, age and testing session or between age, diagnosis, and testing session. The means and SDs for men and women with schizophrenia are shown in online supplementary table S7.

Age

Older persons, regardless of diagnosis, performed more poorly on associative recognition ($b = -.007$, $t(219) = -5.22$, $P < .001$). The correlation between age and task performance was of a medium effect size and was significantly higher for associative recognition ($r(223) = -.37$, $P < .001$) than for either item recognition conditions ($r(223) = -.10$,

$P > .10$, $t(220) = -4.17$, $P < .001$; $r(223) = -.10$, $P > .10$, $t(220) = -4.45$, $P < .001$). The only other main effect for age was that JOVI accuracy diminished with age ($b = -0.002$, $t(222) = -3.12$, $P = .002$) regardless of group membership. There were no significant interactions between age and diagnosis, age, and testing, or between age, diagnosis, and testing.

Reliability of Individual Scores

The results reported thus far are at the level of group averages. We also evaluated stability of scores at the individual level with intraclass correlation coefficients (ICCs) for the agreement among scores between pairs of testing occasions and among all 3. This form of ICC takes into account as sources of error (unreliability) changes in score level (practice) as well as changes in relative position in the group and so is a more accurate estimate of the stability of individuals' scores than a correlation coefficient. Table 4 summarizes the ICCs for patients separately for each test interval pair and also provides r values for comparison to other publications. Of the 64 reliability coefficients computed, only 3 differed significantly between patients and controls. The stability of individual differences in scores over the 3 occasions (table 3, last column) ranged from low (JOVI Threshold = .50) to near adequate (AX-CPT = .69) or adequate (DPX = .74; JOVI proportion correct = .72). The CCE and RiSE measures had moderate ICC stability coefficients (.55–.63). The stability of individual differences from T2 to T3 (the sessions with little to no practice effects) was moderate

Table 4. Test-Retest Reliability for Patients

Task	T1–T2		T2–T3		T1–T3		T1–T2–T3
	ICC	r	ICC	r	ICC	r	ICC
DPX	**.76**	**.78**	**.77**	**.78**	**.70**	**.73**	**.74**
AX-CPT	.67	**.70**	.80	**.80**	.6	.64	.69
CCE difference	.65	.65	.57	.61	.49	.50	.58
JOVI							
Threshold	.47	.50	.56	.56	.45	.47	.50
Accuracy	**.75**	**.78**	**.73**	**.74**	.67	**.73**	**.72**
RiSE							
Associative recognition	.49	.49	.61	.62	.54	.56	.55
Item recognition–associative encoding	.68	.69	.55	.55	.51	.52	.58
Item recognition–item encoding	**.71**	**.72**	.62	.63	.55	.58	.63

Note: Items given in bold font exceed the 0.70 "acceptability for clinical trials" suggested by the MATRICS Consensus Battery Committee.[36]

for the CCE, the RiSE measures, and JOVI threshold, adequate for the DPX, JOVI accuracy, and good for the AX-CPT (eg, .80).

Tolerability Ratings

Tolerability ratings made by the participants are shown in table 5 for each of the 3 testing sessions. The mean tolerability ratings for the MATRICS battery tasks made by participants with schizophrenia ranged from a low of 3.7 to a high of 5.6.[36] For the JOVI, CCE, and RiSE, the lowest mean tolerability rating was 3.9 (JOVI at Time 1) and the highest was 4.8 (RiSE at Time 1). Thus, the CNTRaCS tasks tolerability ratings were overall slightly lower than those for the MATRICS tasks, but in the same range and very similar to the MATRICS computerized CPT (*M* = 4.2).[36]

Discussion

The present results replicate our prior findings of group differences on the 4 CNTRaCS consortium tasks. The DPX, AX-CPT, and RiSE are sensitive measures of specific cognitive impairments observed in persons with schizophrenia that yield large effect sizes for group differences. The JOVI accuracy scores yield a moderate effect size for group differences in this stable chronic outpatient population. However, this measure appears to be state sensitive because all prior studies of this issue indicate greater impairment in samples of more symptomatic, chronic, and/or lower functioning patients.[40–45] Further, as in previous work, the memory impairment of persons with schizophrenia was particularly pronounced for item recognition following relational encoding in comparison with item encoding. Consistent with our earlier report,[27] the CCE did not reveal a reliable deficit in early visual processing aspects of surround suppression

Table 5. Tolerability Ratings for JOVI, CCE, and RiSE

	Mean	Mode	Median
Time 1			
Controls			
CCE			
No Surround	4.3	4	4
Surround	4.0	4	4
JOVI	3.4	3	4
RiSE	5.1	6	5
Patients			
CCE			
No Surround	4.6	4	5
Surround	4.3	4	4
JOVI	3.9	4	4
RiSE	4.8	5	5
Time 2			
Controls			
CCE			
No Surround	4.3	4	4
Surround	4.0	4	4
JOVI	3.4	3	3
RiSE	5.2	6	5
Patients			
CCE			
No Surround	4.8	4	5
Surround	4.3	4	4
JOVI	4.1	4	4
Rise	4.6	5	5
Time 3			
Controls			
CCE			
No Surround	4.3	4	4
Surround	4.2	4	4
JOVI	3.5	3	3
RiSE	5.2	5	5
Patients			
CCE			
No Surround	4.7	4	5
Surround	4.3	4	4
JOVI	4.1	4	4
RiSE	4.7	4	4

(gain control) in our samples of stable schizophrenia outpatients with relatively low levels of positive and disorganization symptoms.

Most importantly for the goals of this work, the psychometric properties of a number of measures make them nicely suitable for both group differences studies and treatment studies that examine changes over time. The evidence was clearest in this regard for the DPX/AX-CPT tasks. The developers of the MATRICS argued that "The committee considered an r value of 0.70 to be acceptable test-retest reliability for clinical trials (p. 5)."[36] The DPX and AX-CPT measures both exceeded this criterion. The distributions were reasonable for the measures from the AX-CPT/DPX, with relatively low skew and relatively few patients at ceiling. Thus, either of these measures would be good choices for treatment studies wishing to assess specific changes in cognitive control and goal maintenance. The evidence for the JOVI and the RiSE was more mixed. One of the JOVI measures, accuracy, had acceptable levels of reliability for use in either group difference or treatment studies, had relatively low levels of skew, and few patients at ceiling. The tolerability ratings for this task were somewhat lower than for the other tasks, which likely reflects the fact that it is a bit challenging visually. In contrast, the threshold measures from the JOVI did not have acceptable reliability for treatment studies, and were more highly skewed. Thus, investigators wishing to assess visual integration using the JOVI in treatment studies (or group difference studies) may wish to focus on the accuracy measures as compared to the threshold measures.

The evidence for the RiSE was also mixed. The Item Recognition measures had higher reliability than the Associative Recognition measure. For example, the reliabilities of the MATRICS Hopkins Verbal Learning Test and the item recognition indices from the RiSE were comparable (.68 and from .68 to .71, respectively). However, there was somewhat more skew for the RiSE Item Recognition measures though they had very few patients at ceiling. Further, the tolerability ratings for the RiSE measures were the highest of all the tasks. Thus, the choice of whether or not to include the RiSE in a treatment study may depend on an investigators relative weighting of validity and reliability. The RiSE has excellent validity as a measure of relational and item encoding and retrieval and thus has excellent interpretability in terms of understanding mechanisms of pathology and potential mechanisms of change in a treatment study. For some investigators, the importance of this validity will make it worth the larger sample size that would be needed in order to use measures with somewhat lower reliability in a treatment study.

Of particular importance in treatment studies are practice effects between baseline and subsequent assessments. We found practice effects for 3 of 8 dependent variables, JOVI accuracy, and the DPX and AX-CPT d'.

The magnitude of the practice effects was not different between patients and controls. Further, the practice effects on the DPX and AX-CPT increased with age. Important recommendations for treatment studies are therefore for the inclusion of younger patients to minimize practice effects and for baseline assessments prior to intervention.

The results also suggest the crossover treatment designs would not be optimal with some of these measures because of practice effects, particularly between the first and second assessment. Specifically, practice effects were greater between the first and second than between the second and third testings. This suggests that multiple baselines may provide important protection in treatment studies employing these tasks by removing the largest practice effect from the evaluation of drug effect and potentially allowing establishment of task strategies by T2 that can remain stable at subsequent testing sessions. We should note, however, that there are some drawbacks to using multiple baselines, including added time and effort and the possibility of increasing the number of individuals at or close to ceiling. Further, it must be noted that the comparisons between T1–T2 and T2–T3 are confounded by different amounts of time between pairs of assessments as the amount of time between T2 and T3 was substantially longer than the time between T1 and T2, which may well have attenuated the second practice effect. However, given that this would be a typical design in a treatment trial (eg, short time between multiple baselines, much longer time between second baseline and posttreatment effects), these results are likely to be generalizable to the context in which such a design would be used. Nonetheless, a study comparing equal intervals between a first and second session and a second and third session would help clarify the issues of differential practice as a function of multiple testing sessions, especially if the times between evaluations were longer intervals more like those used in treatment trials.

Neither sex nor site contributed to variation in task performance. Our lack of sex effects is different from some previous studies that have reported more severe cognitive deficits in men with schizophrenia. However, the majority of previous studies had relatively small samples, and the absence of sex differences in the current relatively well-powered study suggests that if such sex differences are a characteristic of schizophrenia, they are either small in effect size or dependent on clinical characteristics not present in the current study (eg, acutely ill patients).

We also found several effects of age that did not differ as a function of diagnostic group. Older persons tend to do more poorly on many cognitive tasks,[46] but in this study, age was related to overall performance only for JOVI accuracy and RiSE associative recognition memory retrieval. The relationship between aging and increased loss of associative recognition vs item recognition performance is well established.[47] Likewise, contour integration tasks similar to the JOVI have revealed an adverse effect of aging on contour integration ability.[48,49] Note, our examination of

aging effects was somewhat limited because, as is typical of schizophrenia cognitive treatment studies, we did not include participants >65 years of age. In addition, we found greater practice effects on the AX-CPT and DPX among older participants. These differential practice effects could be related to several potential variables that may mediate age effects that were not considered in this study, such as degree of computer experience and diurnal variation in cognitive performance. Both of these factors have been found to influence cognitive performance[50,51] and may also vary as a function of age. As such, it will be important for future studies to assess the degree to which such factors influence cognitive function in schizophrenia, mediate any age effects, and/or influence responsivity to treatment.

In summary, this study provides further evidence of the utility of measures of 3 CNTRaCS constructs for the study of specific neurocognitive deficits in schizophrenia. These data on psychometric properties indicate the suitability of some of the CNTRaCS measures for repeated assessment in appropriately designed treatment trials. The question of differential learning/practice among 3 testing occasions for several of the tasks requires further study though our results do provide suggestive evidence supporting the use of a repeated baseline approach. The associations between age and performance, though limited, indicate that age may require statistical control in studies in which patients are assigned to groups randomly. In such designs, covariance adjustment is a useful means of eliminating within-group variance associated with age, resulting in more powerful between-group tests. Finally, it will be important for future work to also assess the stability of activation in neural systems that support these cognitive processes.

Supplementary Material

Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Funding

Acknowledgments

References

1. Carter CS, Barch DM. Cognitive neuroscience-based approaches to measuring and improving treatment effects on cognition in schizophrenia: the CNTRICS initiative. *Schizophr Bull*. 2007;33:1131–1137.

2. Barch DM, Carter CS, Arnsten A, et al. Selecting paradigms from cognitive neuroscience for translation into use in clinical trials: proceedings of the third CNTRICS meeting. *Schizophr Bull*. 2009;35:109–114.

3. Gold JM, Barch DM, Carter CS, et al. Clinical, functional, and intertask correlations of measures developed by the Cognitive Neuroscience Test Reliability and Clinical Applications for Schizophrenia Consortium. *Schizophr Bull*. 2012;38:144–152.

4. Falleti MG, Maruff P, Collie A, Darby DG. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *J Clin Exp Neuropsychol*. 2006;28:1095–1112.

5. Goldberg TE, Goldman RS, Burdick KE, et al. Cognitive improvement after treatment with second-generation antipsychotic medications in first-episode schizophrenia: is it a practice effect? *Arch Gen Psychiatry*. 2007;64:1115–1122.

6. Pietrzak RH, Snyder PJ, Maruff P. Use of an acute challenge with d-amphetamine to model cognitive improvement in chronic schizophrenia. *Hum Psychopharmacol*. 2010;25:353–358.

7. Goldstein G, Watson JR. Test-retest reliability of the Halstead-Reitan battery and the WAIS in a neuropsychiatric population. *Clin Neuropsychologist*. 1989;3:265–272.

8. Conway Greig T, Nicholls SS, Wexler BE, Bell MD. Test-retest stability of neuropsychological testing and individual differences in variability in schizophrenia outpatients. *Psychiatry Res*. 2004;129:241–247.

9. Manoach DS, Halpern EF, Kramer TS, et al. Test-retest reliability of a functional MRI working memory paradigm in normal and schizophrenic subjects. *Am J Psychiatry*. 2001;158:955–958.

10. Wilk CM, Gold JM, Bartko JJ, et al. Test-retest stability of the Repeatable Battery for the Assessment of Neuropsychological Status in schizophrenia. *Am J Psychiatry*. 2002;159:838–844.

11. Collie A, Maruff P, Darby DG, McStephen M. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc*. 2003;9:419–428.

12. Gilmore GC, Royer FL, Gruhn JJ. Age differences in symbol-digit substitution task performance. *J Clin Psychol*. 1983;39:114–124.

13. Royer FL, Janowitch L. Performance of process and reactive schizophrenics on a symbol-digit substitution task. *Percept Mot Skills*. 1973;37:63–70.

14. Strauss ME, Wagman AM, Quaid KA. Preparatory interval influences on reaction-time of elderly adults. *J Gerontol*. 1983;38:55–57.

15. Saccuzzo DP. Bridges between schizophrenia and gerontology: generalized or specific deficits? *Psychol Bull*. 1977;84:595–600.

16. Harvey PD, Parrella M, White L, Mohs RC, Davidson M, Davis KL. Convergence of cognitive and adaptive decline in late-life schizophrenia. *Schizophr Res*. 1999;35:77–84.

17. Irani F, Brensinger CM, Richard J, et al. Computerized neurocognitive test performance in schizophrenia: a lifespan analysis. *Am J Geriatr Psychiatry*. 2012;20:41–52.

18. Loewenstein DA, Czaja SJ, Bowie CR, Harvey PD. Age-associated differences in cognitive performance in older

patients with schizophrenia: a comparison with healthy older adults. *Am J Geriatr Psychiatry*. 2012;20:29–40.

19. Jiménez J, Mancini-Marïe A, Mendrek A. The case for not combining men and women in neurocognitive studies for schizophrenia. *Schizophr Res*. 2009;108:293–294.

20. Longenecker J, Dickinson D, Weinberger DR, Elvevåg B. Cognitive differences between men and women: a comparison of patients with schizophrenia and healthy volunteers. *Schizophr Res*. 2010;120:234–235.

21. Seidman LJ, Goldstein JM, Goodman JM, et al. Sex differences in olfactory identification and Wisconsin Card Sorting performance in schizophrenia: relationship to attention and verbal ability. *Biol Psychiatry*. 1997;42:104–115.

22. Weiss EM, Deisenhammer EA, Hinterhuber H, Marksteiner J. Gender differences in cognitive functions. *Fortschr Neurol Psychiatr*. 2005;73:587–595.

23. Leung A, Chue P. Sex differences in schizophrenia, a review of the literature. *Acta Psychiatr Scand Suppl*. 2000;401:3–38.

24. Szymanski S, Lieberman JA, Alvir JM, et al. Gender differences in onset of illness, treatment response, course, and biologic indexes in first-episode schizophrenic patients. *Am J Psychiatry*. 1995;152:698–703.

25. Usall J, Suarez D, Haro JM; SOHO Study Group. Gender differences in response to antipsychotic treatment in outpatients with schizophrenia. *Psychiatry Res*. 2007;153:225–231.

26. Henderson D, Poppe AB, Barch DM, et al. Optimization of a goal maintenance task for use in clinical applications. *Schizophr Bull*. 2012;38:104–113.

27. Barch DM, Carter CS, Dakin SC, et al. The clinical translation of a measure of gain control: the contrast-contrast effect task. *Schizophr Bull*. 2012;38:135–143.

28. Ragland JD, Ranganath C, Barch DM, et al. Relational and Item-Specific Encoding (RISE): task development and psychometric characteristics. *Schizophr Bull*. 2012;38:114–124.

29. Silverstein SM, Keane BP, Barch DM, et al. Optimization and validation of a visual integration test for schizophrenia research. *Schizophr Bull*. 2012;38:125–134.

30. Servan-Schreiber D, Cohen JD, Steingard S. Schizophrenic deficits in the processing of context. A test of a theoretical model. *Arch Gen Psychiatry*. 1996;53:1105–1112.

31. MacDonald AW 3rd, Goghari VM, Hicks BM, Flory JD, Carter CS, Manuck SB. A convergent-divergent approach to context processing, general intellectual functioning, and the genetic liability to schizophrenia. *Neuropsychology*. 2005;19:814–821.

32. Wechsler D. *Wechsler Test of Adult Reading*. San Antonia, TX: The Psychological Corporation; 2001.

33. First MB, Spitzer RL, Miriam G, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. New York, NY: Biometrics Research, New York State Psychiatric Institute; 2002.

34. Ventura J, Green MF, Shaner A, Liberman RP. Training and quality assurance on the Brief Psychiatric Rating Scale: the "drift busters". *Int J Methods Psychiatr Res*. 1993;3:221–226.

35. Overall JE, Gorham DR. The brief psychiatric rating scale. *Psychol Reports*. 1962;10:799.

36. Nuechterlein KH, Green MF, Kern RS, et al. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am J Psychiatry*. 2008;165:203–213.

37. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: L. Erlbaum Associates; 1988.

38. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39:175–191.

39. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 2009;41:1149–1160.

40. Silverstein SM, Keane BP. Perceptual organization impairment in schizophrenia and associated brain mechanisms: review of research from 2005 to 2010. *Schizophr Bull*. 2011;37:690–699.

41. Silverstein SM, Kovács I, Corry R, Valone C. Perceptual organization, the disorganization syndrome, and context processing in chronic schizophrenia. *Schizophr Res*. 2000;43:11–20.

42. Uhlhaas PJ, Phillips WA, Mitchell G, Silverstein SM. Perceptual grouping in disorganized schizophrenia. *Psychiatry Res*. 2006;145:105–117.

43. Uhlhaas PJ, Phillips WA, Silverstein SM. The course and clinical correlates of dysfunctions in visual perceptual organization in schizophrenia during the remission of psychotic symptoms. *Schizophr Res*. 2005;75:183–192.

44. Schenkel LS, Spaulding WD, DiLillo D, Silverstein SM. Histories of childhood maltreatment in schizophrenia: relationships with premorbid functioning, symptomatology, and cognitive deficits. *Schizophr Res*. 2005;76:273–286.

45. Silverstein S, Uhlhaas PJ, Essex B, Halpin S, Schall U, Carr V. Perceptual organization in first episode schizophrenia and ultra-high-risk states. *Schizophr Res*. 2006;83:41–52.

46. Salthouse TA. Selective review of cognitive aging. *J Int Neuropsychol Soc*. 2010;16:754–760.

47. Old SR, Naveh-Benjamin M. Differential effects of age on item and associative measures of memory: a meta-analysis. *Psychol Aging*. 2008;23:104–118.

48. Del Viva MM, Agostini R. Visual spatial integration in the elderly. *Invest Ophthalmol Vis Sci*. 2007;48:2940–2946.

49. Roudaia E, Bennett PJ, Sekuler AB. The effect of aging on contour integration. *Vision Res*. 2008;48:2767–2774.

50. Iverson GL, Brooks BL, Ashton VL, Johnson LG, Gualtieri CT. Does familiarity with computers affect computerized neuropsychological test performance? *J Clin Exp Neuropsychol*. 2009;31:594–604.

51. Schmidt C, Collette F, Cajochen C, Peigneux P. A time to think: circadian rhythms in human cognition. *Cogn Neuropsychol*. 2007;24:755–789.