Schizophrenia Bulletin vol. 38 no. 1 pp. 144–152, 2012
doi:10.1093/schbul/sbr142
Advance Access publication on November 17, 2011

Clinical, Functional, and Intertask Correlations of Measures Developed by the Cognitive Neuroscience Test Reliability and Clinical Applications for Schizophrenia Consortium

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

[1]Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, PO Box 21247, Baltimore, MD 21228; [2]Washington University in St Louis; [3]University of California at Davis; [4]University College London; [5]University of Minnesota; [6]Budapest University of Technology and Economics, Budapest, Hungary; [7]University of Medicine and Dentistry of New Jersey; [8]Case Western Reserve University

*To whom correspondence should be addressed; tel: 410-402-7871, fax: 410-402-7198, e-mail: jgold@mprc.umaryland.edu

The goal of the Cognitive Neuroscience Test Reliability and Clinical Applications for Schizophrenia (CNTRACS) Consortium was to develop measures of discrete cognitive processes, allowing for the interpretation of specific deficits that could be linked to specific neural systems. Here we report on the intertask, clinical, and functional correlates of the 4 tasks that were investigated in large groups of patients with schizophrenia (>100) and healthy controls (>73) at 5 sites across the United States. In both healthy and patient groups, the key dependent measures from the CNTRACS tasks were minimally intercorrelated, suggesting that they are measuring discrete abilities. Correlations were examined between CNTRACS tasks and measures of functional capacity, premorbid IQ, symptom severity, and level of community functioning. Performance on tasks measuring relational memory encoding, goal maintenance, and visual gain control were correlated with premorbid IQ and the former 2 tasks with the functional capacity. Goal maintenance task performance was negatively correlated with negative symptom severity and informant reports of community function. These correlations reflect the relationship of specific abilities with functional outcome. They are somewhat lower than functional outcome correlations observed with conventional neuropsychological tests that confound multiple cognitive and motivational deficits. The measures of visual integration and gain control were not significantly correlated with clinical symptoms or function. These results suggest that the CNTRACS tasks measure discrete cognitive abilities, some of which relate to aspects of functional capacity/outcome in schizophrenia.

Key words: cognitive neuroscience/relational memory/executive control/visual integration

The Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (CNTRICS) initiative was designed to identify critical constructs and promising measurement approaches from the basic cognitive neuroscience literature that, with appropriate modification and psychometric validation, could be useful in the context of clinical trials in schizophrenia.[1] The results of that consensus building process have been reported in several papers.[2–4] The Cognitive Neuroscience Test Reliability and Clinical Applications for Schizophrenia (CNTRACS) Consortium followed from the CNTRICS initiative and was intended to do the task optimization and psychometric analysis needed before nominated measures could be recommended for use in clinical trials. We selected for study 4 of the tasks that were nominated as part of the CNTRICS initiative, spanning aspects of visual perception to higher order cognitive control and episodic memory encoding and retrieval processing as these are important areas of impairment in schizophrenia and span multiple neural systems.[5] These tasks were then tested in a multisite study with large samples of schizophrenia patients and healthy controls. The results for each of these measures were described in the 4 other articles accompanying this article[6–9] (see also table 1). Here, we present an analysis of intertask, clinical, and functional correlates of these 4 tasks. In these analyses, we considered 3 fundamental issues that need to be addressed in order to validate these measures for use in clinical research.

One goal of the present study was to determine the extent to which the 4 tasks used here succeeded in the goal of measuring discrete cognitive functions or whether the overall pattern of group differences substantially reflected shared variance across cognitive domains, with relatively little evidence of domain-specific impairment. This issue is of

Table 1. Task Descriptions and Other Relevant Task Information

Task Name	Construct Task Is Intended to Measure	Brief Description of Task	Website URL	Effect Size of Control vs Schizophrenia Difference	Sample Size for Current Analyses
Dot Probe Expectancy	Goal maintenance in working memory	Participants are asked to respond target to "X" probes but only if they come after a valid "A" cue. Thus, participants must maintain the goal set up by the cue in working memory to respond correctly to probes.	http://cntracs.ucdavis.edu/task/dpx	D' Context: $d = 1.03$	SCZ: 138 CON: 136
Relational and Item-Specific Encoding and Retrieval	Relational and item encoding and retrieval	Participants are presented either with pairs of items and asked to encode them relationally (by deciding whether one item could fit inside the other) or with single items and asked to decide whether they are living or nonliving. Participants are then given item and associative recognition tests.	http://cntracs.ucdavis.edu/task/rise	Item recognition following relational encoding: $d = .84$ Item recognition following item encoding: $d = .62$ Associative recognition: $d = .95$	SCZ: 102 CON: 73
Jittered Orientation Visual Integration	Visual integration	Participants briefly observed a field of elements and decided whether a subset of those elements formed a leftward or rightward pointing shape. Integration difficultly depends on the amount of orientation jitter present in the shape's elements	http://cntracs.ucdavis.edu/task/jovi	Log-transformed threshold: $d = .85$	SCZ: 125 CON: 132
Contrast-Contract Effect Task	Gain control	Participants are asked to match a variable contrast patch to a central patch. When the surround is high contrast, the inner target is perceived to be of lower contrast than when the same target is perceived without a surround.	http://cntracs.ucdavis.edu/task/cce	Change in target contrast perception with surround: $d = .31$	SCZ: 130 CON: 132

Note: SCZ, people with schizophrenia; CON, healthy controls.

concern because it is known that the measures in some other cognitive batteries, such as the one developed as part of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) process,[10] are substantially intercorrelated despite their being thought of as measuring discrete cognitive domains.[12,13] Further, when clinical neuropsychological measures are administered as part of a battery to people with schizophrenia and healthy controls, the overall pattern of group differences substantially reflects shared variance across cognitive domains, with relatively little evidence of domain-specific impairment.[13,14] If performance on the CNTRACS measures also reflects the impact of a common generalized deficit, we would expect that the CNTRACS tasks should be highly intercorrelated, which would undermine the utility of the CNTRACS approach. Alternatively, if the CNTRACS are minimally intercorrelated, and patients demonstrate impairments on each of the tasks, this could be seen as indirect evidence that at least some of the overall cognitive dysfunction observed in schizophrenia can be understood as reflecting the impact of specific deficits in different specific neural/cognitive systems.

A second goal of the present study was to consider the clinical correlates of cognitive measures. Evidence from studies using conventional clinical neuropsychological tests suggests that cognitive performance is generally uncorrelated with the severity of positive symptoms, with small to medium correlations often observed with severity of negative and disorganization symptoms.[15] For comparison purposes, it would be important to determine whether the use of more specific cognitive measures would reveal similar magnitude associations to clinical symptoms as found in prior research or whether the use of the CNTRICS measures would result in either enhanced or reduced sensitivity to symptom severity. Enhanced sensitivity could arise if the more specific measures targeted a cognitive process that was thought to play a central role in the development of a specific symptom and provided a better assay for that process than clinical neuropsychological tests.[16,17] Alternatively, one might predict a reduced relationship between measures of disorganized or negative symptoms and measures of specific cognitive functions if the correlations already in the literature between these illness dimensions reflect shared variance with illness severity or a generalized cognitive deficit.

The third goal of this study was to assess the relationship of the CNTRACS measures with functional outcome and with measures of functional capacity. The fact that clinical neuropsychological measures relate to functional outcome was one of the critical pieces of evidence that led the Food and Drug Administration to give approval of cognition as a treatment target.[18,19] Thus, it is essential to determine whether the CNTRACS measures, which are potentially more sensitive and specific, offer more or less power to detect such effects. Superficially, one might suspect that specific cognitive measures might offer less robust predictions of functional outcome. That is, if complex multifactorial clinical neuropsychological measures relate to complex multidetermined outcomes like vocational performance (and these relationships are often more robust using composite cognitive scores that summarize performance across domains[19]), then it seems likely that more specific process measures would show less overlap with measures of functional outcome. Speculatively, one can also imagine how power could be enhanced. If some specific cognitive processes, by themselves, are critical determinants of outcome, then the use of measures that are more "process pure" should increase power as they provide a greater signal to noise ratio for the critical construct than more complex, polyfactorial measures.

Thus, to address the three goals described above, there were four aims to the current study. The first was to examine intercorrelations among performance on the 4 primary cognitive paradigms being developed as part of the CNTRACS Consortium: the Dot Probe Expectancy Task (DPX), the Relational and Item-Specific Encoding Task (RISE), the Contrast-Contrast Effect Task (CCE) and the Jittered Orientation Visual Integration Task (JOVI). The second aim was to assess the relationship of these CNTRACS measures to assessments of reading ability as an estimate of premorbid IQ and a measure of functional capacity. The third and fourth aims were to examine the relationship between performance on these measures and the severity of clinical symptoms and functional outcome among individuals with schizophrenia.

Methods

Participants

The participants for this study, their recruitment, and the inclusion/exclusion criteria are described in detail in the first article in this set of articles.[6]

Diagnosis and Clinical Assessment

The training of clinical raters and the clinical assessment of participants are described in detail in the first article in this set of articles.[6] We computed 3 subscales from the Brief Psychiatric Rating Scale[20] (BPRS) using the following items: (1) Positive Symptoms (Grandiosity, Suspiciousness, Hallucinations, Unusual Thought Content); (2) Disorganization (Bizarre Behavior, Disorientation, Conceptual Disorganization, Mannerisms and Posturing); and (3) Negative Symptoms (Self-Neglect, Blunted Affect, Emotional Withdrawal, Motor Retardation). Interrater reliability, measured using intraclass correlation coefficients with the individual rater as the unit of measurement, was 0.89 for positive, 0.81 for disorganization, and 0.74 for negative symptoms.

Measures of Community Function

We used the Specific Levels of Functioning Scale (SLOF)[21] to assess community function. The SLOF is a highly reliable and well-validated scale of community function that has been shown to relate to proxy measures of functional performance and to performance on neuropsychological tasks.[22] The SLOF contains 43 items that assess the domains of interpersonal relationships, participation in community activities (using public transportation, shopping, use of telephone, bill paying, etc.), and work skills (ability to be on time, ability to stay on task, level of supervision required to complete skills, etc.). Prior work strongly suggests that it is not sufficient to assess real-world function based only on the report of the patient and that reports from someone who knows the patient well (eg, case manager) are more valid.[23] Thus, we asked an individual with information about the patient's function to fill out the SLOF (family member, caseworker, therapist), in addition to collected reports from the patients themselves. In the analyses presented here, we focus on the informant SLOF reports, which we were able to obtain for 114 patients. The patients with and without SLOF informant reports did not differ significantly on age, personal education, parental education, or BPRS scores (all $p > .10$). In addition, the patients with and without SLOF informant reports did not differ significantly on any of the dependent measures for the CNTRACS task or on the UCSD Performance-Based Skills Assessment-Brief (UPSA-B, described below, all $p > .38$). However, the patients with SLOF Informant reports (mean = 34.5, SD = 9.9) did score higher on the Wechlser Test of Adult Reading (WTAR)[24] than those without (mean = 29.0, SD = 8.8) SLOF informant reports ($t(145) = 2.9$, $p = .005$).

Such ratings scale measures of community function are one important way to assess real-world outcomes in schizophrenia. However, such measures rely on self-reports or informant reports and do not directly assess life skills. Thus, measures of functional capacity that assess simulated life skills are an important additional measure of community function. As such, we also used the well-validated UPSA-B.[25–27] In addition, we also used the WTAR, a measure of single-word decoding, to estimate premorbid IQ.[24]

Testing Sessions

The tasks are briefly described in table 1 which also provides the web site URL that can be used to download the tasks (see also Henderson et al,[6] Barch et al,[7] Silverstein et al,[8] and Ragland et al[9] for task details). The procedures and order of testing are described in detail in the first article in this set of articles.[6]

Data Analysis

For the analyses reported in this article, we focused on the primary outcome measures from each of the 4 tasks. In addition, we focused on the version of each task that was deemed to be the most "optimal" based on the analyses presented in the cases where more than one version of the task was evaluated (see Henderson et al,[6] Barch et al,[7] Silverstein et al,[8] and Ragland et al[9] for discussion of task optimization). Thus, for the RISE, we used the visual object version and focused on recognition D′ scores for items encoded in the Item Encoding condition (D'-IRIE), for items encoded in the Relational Encoding condition (D'-IRRE), and for Relational Recognition (D'-RR). There were 102 people with schizophrenia (SCZ) and 73 healthy controls (CON) with data on this version of the RISE. For the DPX, we focused on D'-Context from the short inter-trial interval version with the distribution of trial types providing the greatest prepotency effects (88 AX trials, 16 AY and BX trials, and 8 BY trials). We had data from 138 SCZ and 136 CON for this version of the DPX. For the JOVI, we focused on the threshold estimates, which could be computed for 125 patients and 132 CON. We made some changes to the JOVI part way through the study[8] (eliminating jitter conditions with little variance and added some additional intermediate jitter conditions) but included data from all subjects to maximize power. For the CCE, we focused on the differences between the average contrast values for the last 10 trials (collapsed across streams) for the no-surround versus the surround conditions for the 100-ms ISI condition. We had data from 130 SCZ and 132 CON for this version of the CCE. Given that the *N* for different tasks differed, and given that some participants did not have SLOF data, we will present the *N* for each analysis below.

The sample sizes were very similar across sites, and there were no major site effects on the results. We started by examining the measures for normality and outliers. All the CNTRACS measures were normally distributed, other than the JOVI threshold score. Box plots computed in SPSS revealed only 3 outlier data points, which were 1 control and 1 patient on the JOVI threshold and 1 patient on the CCE Contrast Difference Score. These values were not included in the analyses presented below. Pearson correlations were used to examine task intercorrelations and clinical correlates. However, the same analyses using nonparametric correlations, to address the non-normality of the JOVI threshold score, revealed essentially the same results. Multiple regression was used to examine the predictors of functional outcome measures.

Results

Table 2 provides the means and SDs for demographic and clinical data for the entire sample, as well as the means and SDs for the WTAR, the UPSA-B, and the clinical measures among in the individuals with schizophrenia. The patients and control did not differ significantly in age, gender, ethnicity, parental education, or

Table 2. Demographic and Clinical Characteristics of Participants

Variable	Healthy Control		Schizophrenia Patient		Group Comparison
	Mean	SD	Mean	SD	
Age (y)	36.7	12.0	39.3	11.5	$t = 1.87$, $p = .06$
Gender (% males)	61		55		$\chi^2 = 1.07$, $p = .18$
Ethnicity (% Caucasian)	54		57		$\chi^2 = 7.2$, $p = .40$
Personal Education (y)	14.8	2.02	13.2	2.2	$t = 6.35$, $p < .001$
Personal SES	38.6	10.3	26.0	10.1	$t = 10.43$, $p < .001$
Father Education	13.0	2.84	13.5	3.58	$t = 1.49$, $p = .14$
Mother Education	13.3	2.52	13.3	2.76	$t = 0.10$, $p = .92$
Parental SES	44.4	12.6	42.8	15.2	$t = 0.94$, $p = .35$
WTAR	38.0	8.1	33.2	9.9	$t = 4.40$, $p < .001$
UPSA-B	87.5	8.8	77.0	13.4	$t = 7.70$, $p < .001$
SLOF Self-Report (mean across items)	4.75	0.18	4.23	0.43	$t = 12.81$, $p < .001$
SLOF Informant Report (mean across items)	NA		4.09	0.59	
BPRS Positive Symptoms (mean across items)	NA		2.19	1.16	
BPRS Negative Symptoms (mean across items)	NA		1.85	0.74	
BPRS Disorganized Symptoms (mean across items)	NA		1.30	0.45	

Note: WTAR, Wechlser Test of Adult Reading; UPSA-B, UCSD Performance-Based Skills Assessment-Brief; SLOF, Specific Levels of Functioning Scale; BPRS, Brief Psychiatric Rating Scale.

parental SES,[28] though they did differ in personal education and SES. In addition, as expected, the controls scored higher on the WTAR (a proxy for premorbid IQ), the UPSA-B (a measure of functional capacity), and on self-reports of function on the SLOF.

Correlations Among the CNTRACS Measures

We began by examining intercorrelations among the dependent measures from the different CNTRACS tasks, separately for patients and controls. As shown in table 3, the three different RISE measures were strongly intercorrelated, especially among the patients with SCZ. However, other than significant correlations between the RISE D'-RR and the CCE Contrast Difference score in SCZ, and the JOVI threshold score in CON, there were no other significant intercorrelations among the CNTRACS measures.

Correlations Between the CNTRACS Measures and Premorbid IQ, Functional Capacity, and Key Demographic Characteristics

Next we examined performance between the CNTRACS measures, the WTAR (a measure of premorbid IQ), the UPSA-B (a measure of functional capacity), and demographic variables that could influence general cognitive ability, such as personal education and parental SES. As shown in table 4, the RISE and DPX measures show consistent relationships with both the WTAR and UPSA-B among SCZ, though only the DPX shows a relationship to WTAR and UPSA-B among controls. However, there are relatively few significant correlations of the RISE and DPX measures to either participant education or parental SES among either CON or SCZ (only RISE Relational in CON). The JOVI measure did not show any correlations with the WTAR or the UPSA-B among either SCZ or CON but did show a correlation with parental SES, though only among CON. The CCE Contrast difference score was positively correlated with WTAR and participant education among SCZ but was negatively correlated with parental SES among CON. The WTAR and the UPSA-B were significantly positively correlated, and both the UPSA-B and the WTAR showed a strong relationship to participant education among SCZ (only for the WTAR in CON), as well as significant correlations with parental SES among SCZ, though not CON.

Correlations Between the CNTRACS Measures, Clinical Symptoms, and Community Function

As shown in table 5, consistent with prior results using different versions of the context CPT[29,30] the DPX shows significant, albeit modest, negative correlations with BPRS negative symptoms. None of the other CNTRACS measures showed any significant correlations with

Table 3. Intercorrelations Among CNTRACS Task Measures

	RISE D'-IRIE	RISE D'-IRRE	RISE D'-RR	DPX-D'-Context	JOVI Threshold	CCE Contrast Difference
RISE D'-IRIE	—	***.94** (102)*	***.55** (102)*	*.04 (97)*	*.14 (88)*	*.04 (91)*
RISE D'-IRRE	**.90** (73)**	—	***.61** (102)*	*.07 (97)*	*.18 (88)*	*.13 (91)*
RISE D'-RR	.20 (73)	**.26* (73)**	—	*.08 (97)*	*.19 (88)*	***.21* (91)***
DPX-D'-Context	−.06 (73)	.01 (73)	.14 (73)	—	*.08 (118)*	*.05 (123)*
JOVI Threshold	**.25* (70)**	.23 (70)	.17 (70)	−.11 (131)	—	.12 (113)
CCE Contrast Difference	.05 (70)	.01 (70)	−.07 (70)	−.05 (130)	.07 (129)	—

Note: RISE, Relational and Item-Specific Encoding Task; RISE D'-IRIE, recognition D' for items encoded in the Item Encoding condition; RISE D'-IRRE, recognition D' for items encoded in the Relational Encoding condition; RISE D'-RR, recognition D' for Relational Recognition; DPX, Dot Probe Expectancy Task; JOVI, Jittered Orientation Visual Integration Task; CCE, Contrast-Contrast Effect Task. Correlations for patients are shown in italics above the diagonal, and correlations for controls are shown below the diagonal. Bold values are statistically significant.
*$p < .05$, **$p < .01$.

symptoms. The WTAR, a measure of premorbid IQ, also did not show any significant correlations with clinical symptoms. Performance on the UPSA-B showed a significant negative correlation with negative symptoms.

As shown in table 5, the three measures that showed a relationship to SLOF informant reports of community function were the DPX, the WTAR, and the UPSA-B. However, as described above, all 3 of these measures were intercorrelated, and it would be important to know whether they were accounting for common or unique variance in SLOF informant reports. Thus, we conducted a linear regression in which all three measures were used to predict SLOF informant report scores. This regression was significant ($F(3,104) = 5.2$, $p < .001$) and accounted for ~13% of the variance (adjusted $R^2 = .133$). The UPSA-B had a significant beta weight ($\beta = .26$, $p < .01$), but the WTAR ($\beta = .10$, $p = .35$) and the D'-Context from the DPX ($\beta = .12$, $p = .25$) failed to achieve significance as additional predictors. We computed a similar analysis using D'-Context and UPSA-B scores to predict negative symptoms because both showed significant zero-order correlations. This regression was also significant ($F(2,133) = 6.6$, $p < .001$) and accounted for ~9% of the variance (adjusted $R^2 = .092$). However, for this analysis, both D'-Context from the DPX ($\beta = -.21$, $p < .05$) and the UPSA-B ($\beta = -.18$, $p < .05$) had a significant beta weights.

Discussion

These results establish several clear findings and raise a number of important questions. First, and most importantly, it is clear that this group of tasks demonstrate minimal intercorrelation among themselves and very modest

Table 4. Correlations Between CNTRACS Task Measures, Premorbid IQ, Measures of Functional Capacity, and Demographic Variables

	WTAR		UPSA-B		Participant Education		Parental SES	
	SCZ	CON	SCZ	CON	SCZ	CON	SCZ	CON
RISE D'-IRIE	**.25* (102)**	.01 (73)	**.28** (100)**	−.03 (72)	.07 (102)	.07 (73)	.09 (87)	.04 (69)
RISE D'-IRRE	**.32** (102)**	.06 (73)	**.34** (100)**	−.02 (72)	.18 (102)	.13 (73)	.14 (87)	.04 (69)
RISE D'-RR	**.29** (102)**	**.25* (73)**	**.21* (100)**	.01 (72)	.16 (102)	**23* (72)**	.18 (87)	**.37** (69)**
DPX-D'-Context	**.26** (138)**	**.18* (136)**	**.20** (135)**	**.27* (135)**	.16 (137)	.03 (136)	.13 (116)	.05 (129)
JOVI Threshold	.11 (124)	.02 (131)	.13 (123)	−.06 (131)	.11 (123)	.11 (131)	.10 (108)	**.29** (125)**
CCE Contrast Difference	**.18* (131)**	−.10 (130)	.12 (131)	−.13 (130)	**.26** (130)**	−.02 (130)	.10 (123)	**−.21*(129)**
WTAR	—	—	**.38** (144)**	**.23* (135)**	**.55** (146)**	**.34** (136)**	**.36** (124)**	.14 (129)
UPSA-B	—	—	—	—	**.39** (143)**	.15 (135)	**.26* (124)**	.06 (129)

Note: WTAR, Wechlser Test of Adult Reading; UPSA-B, UCSD Performance-Based Skills Assessment-Brief. Other abbreviations are explained in the first footnote to table 1 and table 3. Bold values are statistically significant.
*$p < .05$, **$p < .01$.

Table 5. Correlations Between CNTRACS Task Measures, Functional Outcome, and Clinical Symptoms

	SLOF Patient Reports	SLOF Informant Reports	BPRS Positive Symptoms	BPRS Disorganization Symptoms	BPRS Negative Symptoms
RISE D'-IRIE	.08 (102)	.07 (78)	.09 (102)	.02 (102)	−.07 (102)
RISE D'-IRRE	.14 (102)	.09 (78)	.11 (102)	−.01 (102)	−.10 (102)
RISE D'-RR	.14 (102)	.08 (78)	.02 (102)	−.10 (102)	.01 (102)
DPX-D'-Context	.10 (140)	**.20* (107)**	−.01 (138)	−.14 (138)	**−.23* (138)**
JOVI Threshold	.16 (124)	.04 (94)	−.08 (124)	−.01 (124)	.02 (124)
CCE Contrast Difference	−.03 (131)	−.07 (103)	.09 (130)	−.04 (130)	−.03 (130)
WTAR	.11 (146)	**.23 (114)**	.07 (146)	.03 (146)	−.03 (146)
UPSA-B	**.18* (143)**	**.39** (112)**	.04 (143)	−.15 (143)	**−.23* (143)**

Note: WTAR, Wechsler Test of Adult Reading; UPSA-B, UCSD Performance-Based Skills Assessment-Brief; SLOF, Specific Levels of Functioning Scale; BPRS, Brief Psychiatric Rating Scale. Other abbreviations are explained in the first footnote to table 3. Correlations for patients are above the diagonal, and correlations for controls are below the diagonal.
$^{*}p < .05$, $^{**}p < .01$.

correlation with the WTAR, a measure of premorbid intelligence. Thus, to the extent that they are reliable indicators, these tasks appear to be measuring fairly discrete cognitive processes. Further, to the degree that the tasks demand or share some more general features/processes (such as the ability to sustain performance over time, comprehend instructions, engage with challenging tasks, etc), such shared processes must not be a major contributor to variance in the observed individual differences in performance in the current study, given the modest level of intercorrelation among the tasks. As a point of contrast, a meta-analysis of the bivariate correlations between the types of tasks included in the MATRICS battery yielded a mean correlation of $r = .37$.[10] Thus, the magnitude of the intercorrelations among tasks observed with the CNTRACS measures are clearly lower than those observed with more standard clinical neuropsychological measures, again consistent with the hypothesis that the CNTRACS measures are assessing more discrete cognitive processes.

The facts that the tasks are only minimally intercorrelated, but that the patient group demonstrates impairment relative to controls on each task leads to a nearly inescapable conclusion: patients are impaired on each task for a different reason. That is—doing poorly on one task does not lead to a prediction of poor performance on the other tasks. This pattern of results cannot be easily explained by a generalized cognitive or motivational deficit without asserting that the tasks are basically yielding nearly pure measurement noise. That is, one could observe this pattern of results if the tasks were all highly unreliable, a possibility that is unlikely given the reliability data presented in the prior papers for each of the tasks.[6–9]

It is noteworthy that the RISE and DPX demonstrated minimal correlations with one another, even though both are thought to depend, in part, on the DLPFC. In the RISE, relational processing is thought to depend on higher order strategic encoding processes that require the DLPFC as shown in previous imaging studies.[31] In the DPX, the ability to maintain context representations and task rules also has been shown to demand DLPFC processing.[32,33] However, this DLPFC activity occurs as part of activity in larger neural networks that likely differs substantially across the tasks, perhaps explaining the lack of observed correlation.

More challenging to consider is the fact that just as the tasks fail to correlate with each other at levels expected of clinical neuropsychological tests, they also correlate at lower levels with the UPSA and SLOF than has been reported with clinical neuropsychological tasks.[34,35] It is possible that this reflects biases in our sampling of the population of individuals with schizophrenia. One rarely seeks to recruit the most severely impaired patients to participate in challenging cognitive or imaging protocols and when one does, the rate of refusal is nontrivial. Similarly, given that we did not study inpatients, and only included clinically stable, nonrecreation drug using outpatients, our sample also includes a lower proportion of dual-diagnosis and severely symptomatic patients than many reported in the literature. Thus, it is possible that our study's volunteer population was somewhat less symptomatic and higher functioning than some other study populations reported in the literature. This hypothesis is consistent with the fact that the correlation between the UPSA-B and the SLOF was also somewhat lower than found in a number of prior studies.[36] This restriction of range may also have contributed to the relatively modest/minimal relationships we observed with BPRS symptom domain scores, with the exception of the DPX, where our findings were consistent with prior observations.[29,30]

Note, however, that the UPSA-B performance of our sample closely approximates that reported by Harvey et al[37] in the VALERO study, suggesting that our study population is fairly representative of the types of patients who participate in these types of studies.

The relatively modest correlations between the CNTRACS measures and functional outcome/symptoms, somewhat lower than observed with some standard neuropsychological tests, could—in principle—be the result of the CNTRACS measures having lower reliability than standard neuropsychological tests. The four other articles on each of the individual tasks report on measures of internal consistency that suggest at least reasonable reliability for the CNTRACS measures, reducing the likelihood that our findings are explained on this basis.[6–9] However, the results of the ongoing study on test–retest reliability will be necessary in order to more fully compare the reliability of the CNTRACS measures to measures from neuropsychological batteries. Lastly, it is also possible that the use of SLOF informant reports presented some challenges to assessing functional outcome. This is because the nature and quality of the informant was variable across patients and sites, with some patients having informants with whom they had close contact (who could presumably provide highly valid assessments) and some patients having informants with whom they only had minimal contact (eg, case managers who only saw them once or a month or less) and may have provided less valid reports. Again, the fact that the correlation between the UPSA-B and SLOF informant reports was also somewhat lower than in some prior studies would be consistent with this hypothesis. To address this question, the ongoing test–retest reliability study is including an additional measure of functional capacity, the Multidimensional Scale of Independent Functioning,[38] that uses a standardized set of probes and detailed rating anchors that may offer a more uniform assessment of functional status across patients.

Even allowing for these sampling, reliability, and measurement issues, it is also possible that the relatively modest correlations with functional outcome and functional capacity measures is a real signal, suggesting that assessments of more specific cognitive processes may come at the cost of reduced relationships to clinical outcomes of interest. As noted previously, this type of attenuated relationship makes theoretical sense—more complex multicomponent measures are more likely to be strongly related to performing complex tasks (as may be required in many forms of employment and in managing independent living) than focal measures of a single process. Alternatively, the correlations observed between the CNTRACS measures and functional outcome allow for enhanced interpretive precision of the cognitive process that is related to everyday performance in the community. If replicated, this finding may have important implications for how CNTRACS measures are used in clinical trials. That is, if the FDA approval pathway requires the use of measures with a demonstrated relationship to functional outcomes, then the MATRICS approach may be preferable to the CNTRACS approach. Alternatively, the CNTRACS measures may offer greater precision and be more useful than the MATRICS battery in proof of concept and early phase 2 studies. That is, performance on construct pure measures may specify what a drug does at the cognitive and neural level of analysis, whereas drug effects on the MATRICS battery may assess the impact of that change on more general aspects of cognition that are more strongly correlated with everyday functional performance. Both types of information are valuable but address different questions. Nonetheless, an additional large-scale replication study is underway that includes multiple measures of functional outcome and will include a subset of tasks from the MATRICs battery, allowing a direct comparison of the relationships between cognitive performance, functional outcome, and clinical symptoms with both the CNTRACS tasks and more traditional neuropsychological measures. Regardless of the outcome, it is likely that both types of cognitive measures will play important, though potentially different, roles in the cognitive treatment development process in schizophrenia.

Funding

National Institute of Mental Health (5R01MH084840-03 to D.M.B., 5R01MH084826-03 to C.S.C., 5R01MH084828-03 to S.M.S., 5R01MH084821-03 to J.M.G., and 5R01MH084861-03 to A.W.M.).

Acknowledgments

We acknowledge the hard work and dedication of staff at each of the CNTRACs sites. We also thank our participant for their essential contributions to this study. Dr Gold has received grants from National Institutes of Health, receives royalty payments from the Brief Assessment of Cognition in Schizophrenia, and has consulted with Pfizer, Merck, Astra Zenaca, Solvay, and Glaxo Smith Kline. Dr Barch has received grants from the National Institute of Mental Health (NIMH), National Institute of Aging, National Alliance for Research on Schizophrenia and Depression (NARSAD), Allon, Novartis, and the McDonnell Center for Systems Neuroscience. Dr Carter has received research grants from the NIMH, National Institute on Drug Abuse, the Robert Wood Johnson Foundation and from Glaxo Smith Kline and has been an external consultant for Roche, Servier, Lilly, Merck and Pfizer. Dr MacDonald has received research grants from the NIH and NARSAD. Dr Ragland has received research grants from the NIH and NARSAD. Dr Silverstein has received research grants from NIH, Pfizer, and AstraZeneca. Dr Strauss has no currently active grant or contract support from private or public sources.

References

1. Carter CS, Barch DM. Cognitive neuroscience-based approaches to measuring and improving treatment on cognition in schizophrenia: the CNTRICS initiative. *Schizophr Bull*. 2007;33:1131–1137.
2. Ragland JD, Cools R, Frank M, et al. CNTRICS final task selection: long-term memory. *Schizophr Bull.* 2009;35:197–212.
3. Green MF, Butler PD, Chen Y, et al. Perception measurement in clinical trials of schizophrenia: promising paradigms from CNTRICS. *Schizophr Bull.* 2009;35:163–81.
4. Barch DM, Braver TS, Carter CS, Poldrack RA, Robbins TW. CNTRICS final task selection: executive control. *Schizophr Bull.* 2009;35:115–35.
5. 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;35:109–14.
6. Henderson D, Poppe A, Barch DM, et al. Optimization of a goal maintenance task for use in clinical applications. *Schizophr Bull*. 2012;38:104–113.
7. 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.
8. Silverstein SM, Keane BP, Barch DM, et al. Optimization and validation of a measure of visual Integration for use in schizophrenia treatment research. *Schizophr Bull*. 2012;38:125–134.
9. 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.
10. Dickinson D, Gold JM. Less unique variance than meets the eye: Overlap among traditional neuropsychological dimensions in schizophrenia. *Schizophrenia Bulletin*. 2008;34:423–434.
11. Nuechterlein KH, Barch DM, Gold JM, Goldberg TE, Green MF, Heaton RK. Identification of separable cognitive factors in schizophrenia. *Schizophrenia Research*. 2004;72:29–39.
12. 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.
13. Dickinson D, Iannone VN, Wilk CM, Gold JM. General and specific cognitive deficits in schizophrenia. *Biol Psychiatry*. 2004;55:826–833.
14. Dickinson D, Ragland JD, Gold JM, Gur RC. General and specific cognitive deficits in schizophrenia: Goliath defeats David? *Biol Psychiatry*. 2008;64:823–827.
15. Dominguez Mde G, Viechtbauer W, Simons CJ, van Os J, Krabbendam L. Are psychotic psychopathology and neurocognition orthogonal? A systematic review of their associations. *Psychological Bull.* 2009;135:157–171.
16. Kerns JG, Berenbaum H, Barch DM, Banich MT, Stolar N. Word production and its relationship to positive symptoms. *Psychiatry Res*. 1999;87:29–37.
17. Uhlhass PJ, Silverstein SM. Perceptual organization in schizophrenia spectrum disorders: empirical research and theoretical implications. *Psychol Bull*. 2005;131:618–632.
18. Green MF. What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry*. 1996;153:321–330.
19. Green MF, Kern RS, Braff DL, Mintz J. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the ''right stuff''. *Schizophr Bull*. 2000;26:119–136.
20. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep*. 1962;10:799.
21. Schneider LC, Struening EL. SLOF: a behavioral rating scale for assessing the mentally ill. *Soc Work Res Abstracts*. 1983;19:9–21.
22. Bowie CR, Leung WW, Reichenberg A, et al. Predicting schizophrenia patients real-world behavior with specific neuropsychological and functional capacity measures. *Biol Psychiatry*. 2008;63:505–511.
23. Bowie CR, Twamley EW, Anderson H, Halpern B, Patterson TL, Harvey PD. Self-assessment of functional status in schizophrenia. *J Psychiatr Res*. 2007;41:1012–1018.
24. Wechsler D. *Wechsler Test of Adult Reading (WTAR)*. San Antonio, TX: The Psychological Corporation; 2001.
25. Patterson TL, Goldman S, McKibbin CL, Hughs T, Jeste DV. UCSD Performance-Based Skills Assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophr Bull*. 2001;27:235–245.
26. Twamley EW, Doshi RR, Nayak GV, et al. Generalized cognitive impairments, ability to perform everyday tasks, and level of independence in community living situations of older patients with psychosis. *Am J Psychiatry*. 2002;159:2013–2020.
27. Harvey PD, Velligan DI, Bellack AS. Performance-based measures of functional skills: usefulness in clinical treatment studies. *Schizophr Bull*. 007;33:1138–1148.
28. Hollingshead AD, Redlich FC. *Social Class and Mental Illness.* New York: Wiley; 1958.
29. Cohen JD, Barch DM, Carter CS, Servan-Schreiber D. Schizophrenic deficits in the processing of context: converging evidence from three theoretically motivated cognitive tasks. *J Abnorm Psychol*. 1999;108:120–133.
30. Barch DM, Carter CS, MacDonald A, Braver TS, Cohen JD. Context processing deficits in schizophrenia: diagnostic specificity, longitudinal course, and relationships to clinical symptoms. *J Abnorm Psychol*. 2003;112:132–143.
31. Murray LJ, Ranganath C. The dorsolateral prefrontal cortex contributes to successful relational memory encoding. *J Neurosci*. 2007;27:5515–5522.
32. MacDonald AW, III. Building a clinically relevant cognitive task: case study of the AX paradigm. *Schizophr Bull*. 2008;34:619–628.
33. MacDonald AW, III, Carter CS, Kerns JG, et al. Specificity of prefrontal dysfunction and context processing deficits to schizophrenia in never-medicated patients with first-episode psychosis. *Am J Psychiatry*. 2005;162:475–484.
34. Keefe RS, Fox KH, Harvey PD, Cucchiaro J, Siu C, Loebel A. Characteristics of the MATRICS Consensus Cognitive Battery in a 29-site antipsychotic schizophrenia clinical trial. *Schizophr Res*. 2010;125:161–168.
35. Harvey PD, Helldin L, Bowie CR, et al. Performance-based measurement of functional disability in schizophrenia: a cross-national study in the United States and Sweden. *Am J Psychiatry*. 2009;166:821–827.
36. Bowie CR, Depp C, McGrath JA, et al. Prediction of real-world functional disability in chronic mental disorders: a comparison of schizophrenia and bipolar disorder. *Am J Psychiatry*. 2010;167:1116–1124.
37. Harvey PD, Raykov T, Twamley EW, Vella L, Heaton RK, Patterson TL. Validating the measurement of real-world functional outcomes: phase 1 results of the VALERO study [published online ahead of print May 15, 2011]. *Am J Psychiatry*. doi: 10.1176/appi.ajp.2011.10121723.
38. Jaeger J, Berns SM, Czobor P. The multidimensional scale of independent functioning: a new instrument for measuring functional disability in psychiatric populations. *Schizophr Bull.* 2003;29:153–168.