The MATRICS Consensus Cognitive Battery, Part 1: Test Selection, Reliability, and Validity

Keith H. Nuechterlein, Ph.D.

Michael F. Green, Ph.D.

Robert S. Kern, Ph.D.

Lyle E. Baade, Ph.D.

Deanna M. Barch, Ph.D.

Jonathan D. Cohen, M.D., Ph.D.

Susan Essock, Ph.D.

Wayne S. Fenton, M.D.

Frederick J. Frese III, Ph.D.

James M. Gold, Ph.D.

Terry Goldberg, Ph.D.

Robert K. Heaton, Ph.D.

Richard S.E. Keefe, Ph.D.

Helena Kraemer, Ph.D.

Raquelle Mesholam-Gately, Ph.D.

Larry J. Seidman, Ph.D.

Ellen Stover, Ph.D.

Daniel R. Weinberger, M.D.

Alexander S. Young, M.D., M.S.H.S.

Steven Zalcman, M.D.

Stephen R. Marder, M.D.

Objective: The lack of an accepted standard for measuring cognitive change in schizophrenia has been a major obstacle to regulatory approval of cognition-enhancing treatments. A primary mandate of the National Institute of Mental Health's Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative was to develop a consensus cognitive battery for clinical trials of cognition-enhancing treatments for schizophrenia through a broadly based scientific evaluation of measures.

Method: The MATRICS Neurocognition Committee evaluated more than 90 tests in seven cognitive domains to identify the 36 most promising measures. A separate expert panel evaluated the degree to which each test met specific selection criteria. Twenty tests were selected as a beta battery. The beta battery was administered to 176 individuals with schizophrenia and readministered to 167 of them 4 weeks later so that the 20 tests could be compared directly.

Results: The expert panel ratings are presented for the initially selected 36 tests. For the beta battery tests, data on test-retest reliability, practice effects, relationships to functional status, practicality, and tolerability are presented. Based on these data, 10 tests were selected to represent seven cognitive domains in the MATRICS Consensus Cognitive Battery.

Conclusions: The structured consensus method was a feasible and fair mechanism for choosing candidate tests, and direct comparison of beta battery tests in a common sample allowed selection of a final consensus battery. The MATRICS Consensus Cognitive Battery is expected to be the standard tool for assessing cognitive change in clinical trials of cognition-enhancing drugs for schizophrenia. It may also aid evaluation of cognitive remediation strategies.

(Am J Psychiatry Nuechterlein et al.; AiA:1–11)

Despite the importance of cognitive deficits in schizophrenia, no drug has been approved for treatment of this aspect of the illness. The absence of a consensus cognitive battery has been a major impediment to standardized evaluation of new treatments to improve cognition in this disorder (1, 2). One of the primary goals of the National Institute of Mental Health's (NIMH's) Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative was to develop a consensus cognitive battery for use in clinical trials in schizophrenia. The development of a standard cognitive battery through a consensus of experts was designed to establish an accepted way to evaluate cognition-enhancing agents, thereby providing a pathway for approval of such new medications by the U.S. Food and Drug Administration (FDA). It would also aid in standardized evaluation of other interventions to treat the core cognitive deficits of schizophrenia.

The desirable characteristics of the battery were determined through an initial survey of 68 experts (3). The MATRICS Neurocognition Committee then reviewed and integrated results from all available factor-analytic studies of cognitive performance in schizophrenia to derive separable cognitive domains (4). An initial MATRICS consensus conference involving more than 130 scientists from academia, government, and the pharmaceutical industry led to agreement on seven cognitive domains for the battery and on five criteria for test selection (1). The criteria emphasized characteristics required for cognitive measures in the context of clinical trials: test-retest reliability; utility as a repeated measure; relationship to functional status; potential changeability in response to pharmacological

agents; and practicality for clinical trials and tolerability for patients. The seven cognitive domains included six from multiple factor-analytic studies of cognitive performance in schizophrenia—speed of processing; attention/vigilance; working memory; verbal learning; visual learning; and reasoning and problem solving (4). The seventh domain, social cognition, was included because it was viewed as an ecologically important domain of cognitive deficit in schizophrenia that shows promise as a mediator of neurocognitive effects on functional outcome (5, 6), although studies of this domain in schizophrenia are too new for such measures to have been included in the various factor-analytic studies. Participating scientists initially provided more than 90 nominations of cognitive tests that might be used to measure performance in the seven cognitive domains.

In this article, we describe the procedures and data that the MATRICS Neurocognition Committee employed to select a final battery of 10 tests—the MATRICS Consensus Cognitive Battery (MCCB)—from among the nominated cognitive tests. These procedures involved narrowing the field to six or fewer tests per cognitive domain; creating a database of existing test information; using a structured consensus process involving an interdisciplinary panel of experts to obtain ratings of each test on each selection criterion; selecting the 20 most promising tests for a beta version of the battery; administering the beta battery to individuals with schizophrenia at five sites to directly compare the 20 tests (phase 1 of the MATRICS Psychometric and Standardization Study); and selecting the final battery based on data from this comparison.

A second article in this issue (7) describes the development of normative data for the MCCB using a community sample drawn from the same five sites, stratified by age, gender, and education (phase 2 of the MATRICS Psychometric and Standardization Study). This step was critical to making the consensus battery useful in clinical trials.

During the MATRICS process, it became apparent that the FDA would require that a potential cognition-enhancing agent demonstrate efficacy on a consensus cognitive performance measure as well as on a "coprimary" measure that reflects aspects of daily functioning. Potential coprimary measures were therefore evaluated, as reported in a third article in this issue (8).

Because of space limitations, a brief description of the methods used to evaluate the nominated cognitive measures is presented here; more details are available in a data supplement that accompanies the online edition of this article.

From Test Nominations to a Beta Battery

Summary of Methods and Results

Initial evaluation. The MATRICS Neurocognition Committee, cochaired by Drs. Nuechterlein and Green and including representatives from academia (Drs. Barch, Cohen, Essock, Gold, Heaton, Keefe, and Kraemer), NIMH (Drs. Fenton, Goldberg, Stover, Weinberger, and Zalcman), and consumer advocacy (Dr. Frese), initially evaluated the extent to which the 90 nominated tests met the test selection criteria based on known reliability and validity as well as feasibility for clinical trials. Because the survey established that the battery would optimally not exceed 90 minutes, individual tests with high reliability and validity that took less than 15 minutes were sought. In this initial review, 36 candidate tests across seven cognitive domains were selected.

Expert panel ratings. Procedures based on the RAND/UCLA appropriateness method were used to systematically evaluate the 36 candidate tests (9, 10). This method starts with a review of all relevant scientific evidence and then uses, iteratively, methods that help increase agreement among members of an expert panel that represents key stakeholder groups. A summary of available published and unpublished information on each candidate test was compiled into a database by the MATRICS staff, including information relevant to each test selection criterion (see the Conference 3 database at www.matrics.ucla.edu).

Using this database, an expert panel then evaluated and rated the extent to which each of the 36 candidate tests met each of the five selection criteria. The panel included experts on cognitive deficits in schizophrenia, clinical neuropsychology, clinical trials methodology, cognitive science, neuropharmacology, clinical psychiatry, biostatistics, and psychometrics. These preconference ratings were then examined to identify any that reflected a lack of consensus. Twenty of the 180 ratings indicated a notable lack of consensus, so the expert panel discussed each of these and completed the ratings again. Dispersion decreased, and the median values for nine ratings changed. The median values of all final ratings are presented in Table 1, grouped by cognitive domain.

Selection of beta battery. The Neurocognition Committee used the expert panel ratings to select the beta version of the battery for the MATRICS Psychometric and Standardization Study. The goal was to select two to four measures per domain. The resulting beta version of the MCCB included 20 tests (see Table 2; for more details, see the online data supplement and http://www.matrics.ucla.edu/matrics-psychometrics-frame.htm).

From Beta Battery to Final Battery

Summary of Methods

The MATRICS Psychometric and Standardization Study was conducted to directly compare the tests' psychometric properties, practicality, and tolerability to allow the best representative(s) of each domain to be selected for

TABLE 1. Final Median Ratings by MATRICS Expert Panel for 36 Candidate Cognitive Tests on Five Selection Criteria

Domain and Candidate Tests	Selection Criterion[a]				
	Test-Retest Reliability	Utility as a Repeated Measure	Relationship to Functional Outcome	Potential Changeability in Response to Pharmacological Agents	Practicality and Tolerability
Speed of processing					
Category fluency test, animal naming	7	7.5	5.5	7.5	7.5
Brief Assessment of Cognition in Schizophrenia (BACS), symbol coding subtest	8	5.5	6	6	7
Trail Making Test, Part A	7	7	5.5	6	7.5
Wechsler Adult Intelligence Scale, 3rd ed. (WAIS-III), digit symbol-coding subtest	8	7	6	7	7.5
Symbol Digit Modalities Test	7	6	5	6	7.5
Attention/vigilance					
3–7 Continuous Performance Test	6.5	8	6.5	5	7
Continuous Performance Test—Identical Pairs version	7	5.5	6	5.5	6
A-X Continuous Performance Test, context version	5	7	4	5	5
Working memory					
N-Back (2-Back, NIMH version)	4.5	6	1.5	7	7
N-Back (2-Back, Pittsburgh version)	4	7	1.5	5	6
BACS, digit sequencing subtest	7	7	6.5	5.5	7
Wechsler Memory Scale, 3rd ed. (WMS-III), spatial span subtest	7	8	3.5	5	7
WAIS-III, letter-number sequencing subtest	7	7	6.5	5	7
Spatial delayed response task	4.5	5	5	6.5	5.5
Verbal learning and memory					
WMS-III, logical memory subtest	7	4	6	4.5	6.5
Neuropsychological Assessment Battery, daily living memory subtest	6	6	5	5	7
Neuropsychological Assessment Battery, list learning subtest	7	6	4.5	5	7
California Verbal Learning Test, 2nd ed.	7	6	6	7	7
WMS-III, verbal paired associates subtest	7	5	3	4.5	7
Hopkins Verbal Learning Test—Revised	6	6	4	5	7
Visual learning and memory					
WMS-III, family pictures subtest	6	5	5	4	4.5
Neuropsychological Assessment Battery, shape learning subtest	5	6	6	5	6
Penn Face Memory Test	3.5	3.5	3	4.5	6
WMS-III, visual reproduction subtest	5.5	5.5	5	3.5	5.5
Brief Visuospatial Memory Test—Revised	7	5.5	5	5	7
Visual Object Learning Test	4.5	4	4	5	6
Reasoning and problem solving					
WAIS-III, block design subtest	7	6.5	6	7	6
BACS, Tower of London subtest	6	7	5.5	6	6
Neuropsychological Assessment Battery, mazes subtest	6	7	5.5	6	6.5
Penn Conditional Exclusion Test	6	5	6	6.5	6
Neuropsychological Assessment Battery, categories subtest	6	5.5	4	4.5	6
Delis-Kaplan Executive Functioning Scale, sorting subtest	5	6	4	5	6
Social cognition					
The Awareness of Social Inference Test	5	4	6	4	6
Mayer-Salovey-Caruso Emotional Intelligence Test, managing emotions branch	7	6	6	4	7
Mayer-Salovey-Caruso Emotional Intelligence Test, perceiving emotions branch	7	6	6	5	7
Emotion Recognition–40	3	4	4	4	6

[a] For test-retest reliability, utility as a repeated measure, and practicality and tolerability, MATRICS panelists were asked to rate these tests on a 9-point scale (1=poor, 3=fair, 5=good, 7=very good, 9=superb). For relationship to functional outcome, the anchor points of the 9-point scale were slightly different (1=none, 3=weak, 5=moderate, 7=strong, 9=very strong). For the 9-point scale of potential changeability in response to pharmacological agents, the 1–3 range indicated "unlikely to be sensitive," the 4–6 range indicated "difficult to judge," and the 7–9 range indicated "likely to be sensitive."

TABLE 2. Test-Retest Reliability of the 20 Tests in the MATRICS Beta Battery, by Cognitive Domain

Domain, Tests, and Reference Number	Test Scores Used	Intraclass Correlation Coefficient
Speed of processing		
Category fluency test, animal naming (11)	Total number of animals named in 60 seconds	0.74
Trail Making Test, Part A (12)	Time to completion	0.75
Wechsler Adult Intelligence Scale, 3rd ed. (WAIS-III), digit symbol-coding subtest (13)	Total number correct	0.83
Brief Assessment of Cognition in Schizophrenia (BACS), symbol coding subtest (14)	Total number correct	0.85
Attention/vigilance		
3–7 Continuous Performance Test, shortened version (15)	Overall d'	0.60
Continuous Performance Test—Identical Pairs version (16)	Mean d' value across 2-, 3-, and 4-digit conditions	0.84
Working memory		
BACS, digit sequencing subtest (14)	Number of correct responses	0.75
WAIS-III, letter-number sequencing subtest (13)	Number of correct trials	0.75
Letter-Number Span test (17)	Number of correct trials	0.78
Wechsler Memory Scale, 3rd ed., spatial span subtest (18)	Sum of raw scores on forward and backward conditions	0.74
Spatial delayed response task (19)	Distance from presented dot to remembered dot	0.73
Verbal learning		
Neuropsychological Assessment Battery, daily living memory subtest (20)	Total correct free recall across three trials	0.74
Hopkins Verbal Learning Test—Revised (21), immediate recall	Total number of words recalled correctly over three learning trials	0.68
Visual learning		
Neuropsychological Assessment Battery, shape learning subtest (20)	Total learning score over three trials	0.61
Brief Visuospatial Memory Test—Revised (22)	Total recall score over three learning trials	0.71
Reasoning and problem solving		
WAIS-III, block design subtest (13)	Total raw score	0.84
BACS, Tower of London subtest (14)	Number of correct trials	0.58
Neuropsychological Assessment Battery, mazes subtest (20)	Total raw score	0.83
Social cognition		
Mayer-Salovey-Caruso Emotional Intelligence Test, perceiving emotions branch (23)	Branch score using general consensus scoring	0.80
Mayer-Salovey-Caruso Emotional Intelligence Test, managing emotions branch (23)	Branch score using general consensus scoring	0.73

the final battery. Details of the study's methods are provided in the online data supplement.

Sites and Participants

The study sites had extensive experience with schizophrenia clinical trials and expertise in neuropsychological assessment: University of California, Los Angeles; Duke University, Durham, N.C.; Maryland Psychiatric Research Center, University of Maryland, Baltimore; Massachusetts Mental Health Center, Boston; and University of Kansas, Wichita. Each site contributed at least 30 participants with schizophrenia or schizoaffective disorder, depressed type, who were tested twice, 4 weeks apart.

Study Design and Assessments

Potential participants received a complete description of the study and then provided written informed consent, as approved by the institutional review boards of all study sites and the coordinating site. Next, the Structured Clinical Interview for DSM-IV (24) was administered to each potential participant. If entry criteria were met, baseline assessments were scheduled. Participants were asked to return 4 weeks later for a retest.

In addition to the 20 cognitive performance tests, data collected included information about clinical symptoms (from the Brief Psychiatric Rating Scale [BPRS; 25, 26]), self-report measures of community functioning (from the Birchwood Social Functioning Scale [27] supplemented with the work and school items from the Social Adjustment Scale [28]), measures of functional capacity, and interview-based measures of cognition (8, 29, 30). See the online data supplement for descriptions of alternate cognitive test forms and staff training for neurocognitive assessments, symptom ratings, and community functioning measures.

Results

Participants

Across the five study sites, 176 patients were assessed at baseline, and 167 were assessed again at the 4-week follow-up (a 95% retention rate). Participants' mean age was 44.0 years (SD=11.2), and their mean educational level was 12.4 years (SD=2.4). Three-quarters (76%) of the participants were male. The overall ethnic/racial distribution of the sample was 59% white (N=104), 29% African American (N=51), 6% Hispanic/Latino (N=11), 1% Asian or Pacific Is-

lander (N=2), <1% Native American or Alaskan (N=1), and 4% other (N=7).

Based on the diagnostic interviews, 86% of participants received a diagnosis of schizophrenia and 14% a diagnosis of schizoaffective disorder, depressed type. At assessment, 83% were taking a second-generation antipsychotic, 13% a first-generation antipsychotic, and 1% other psychoactive medications only; current medication type was unknown for 3%. Almost all participants were outpatients, but at one site patients in a residential rehabilitation facility predominated.

As expected for clinically stable patients, symptom levels were low. At the initial assessment, the mean BPRS thinking disturbance factor score was 2.6 (SD=1.3), and the mean BPRS withdrawal-retardation factor score was 2.0 (SD=0.9). Ratings were similar at the 4-week follow-up: the mean thinking disturbance score was 2.4 (SD=1.2), and the mean withdrawal-retardation score was 2.0 (SD=0.8).

Dimensions of Community Functional Status

A principal-components analysis with the seven domain scores from the Social Functioning Scale and a summary measure of work or school functioning from the Social Adjustment Scale yielded a three-factor solution (social functioning, independent living, and work functioning; see supplementary Table 1 in the online data supplement) that explained 59% of the variance and was consistent with previous findings (31, 32). Factor scores from these three outcome domains, as well as a summary score across domains, were used as dependent measures for functional outcome.

Site Effects

Cognitive performance was generally consistent across sites; only four of the 20 analyses of variance (ANOVAs) showed a significant site effect, and the differences were relatively small. In contrast, there were clear site differences in community functioning. ANOVAs revealed significant differences in social outcome (F=3.36, df=4, 170, p<0.02) and independent living (F=4.18, df=4, 170, p<0.01). Work outcome showed a similar tendency (F= 2.35, df=4, 170, p=0.06), and three of the pairwise comparisons were significant.

Test-Retest Reliability

At MATRICS consensus meetings, high test-retest reliability was considered the most important test feature in a clinical trial. Test-retest reliability data are summarized in Table 2. We considered both Pearson's r and the intraclass correlation coefficient, which takes into account changes in mean level (for Pearson's r, see the supplementary Table 2 in the online data supplement). Alternate forms were used for five of the tests. Test-retest reliabilities were generally good. The committee considered an r value of 0.70 to be acceptable test-retest reliability for clinical trials. Most of the tests achieved at least that level.

Utility as a Repeated Measure

Tests were considered useful for clinical trials if they showed relatively small practice effects or, if they had notable practice effects, scores did not approach ceiling performance. We considered performance levels at baseline and 4-week follow-up, as well as change scores, magnitude of change, and the number of test administrations with scores at ceiling or floor. Practice effects were generally quite small (Table 3), but several were statistically significant. Some tests in the speed of processing and the reasoning and problem-solving domains showed small practice effects (roughly one-fifth of a standard deviation), but even so, there were no noticeable ceiling effects or constrictions of variance at the second testing.

Relationship to Self-Reported Functional Outcome

As mentioned above, the sites differed substantially in participants' functional status. At one site, only one participant was working, and another site largely involved patients in a residential rehabilitation program. As a result, correlations between cognitive measures and functional outcome showed considerable variation from site to site. Statistical analyses indicated that the heterogeneity of correlation magnitudes across sites was somewhat greater than would be expected by chance, particularly for the independent living factor. Pooled correlations weighted by sample size, however, were very similar to the overall correlations across sites. Given variability in strength of relationships, the MATRICS Neurocognition Committee examined the correlations by combining participants across sites and by looking at the five sites separately and considering the median correlation among sites. Both methods may underestimate the correlations that might be achieved without such site variations (e.g., nine of the tests had correlations >0.40 with global outcome at one or more sites). Furthermore, the restriction of the functional outcome measures to self-report data may have reduced the correlation magnitudes. However, the data did allow direct comparisons among the tests within the same sample, which was the primary goal (Table 4). The correlations tended to be larger for work outcome and smaller for social outcome, consistent with other studies (33, 34).

Practicality and Tolerability

Practicality refers to the test administrator's perspective. It includes consideration of test setup, staff training, administration, and scoring. In assessing practicality, a 7-point Likert scale was used in three categories (setup, administration, and scoring) and in a global score. Ratings by each tester were made after data collection for the entire sample. Tolerability refers to the participant's view of a test. It can be influenced by the length of the test and any feature making completing the test more or less pleasant, including an unusual degree of difficulty or excessive repetitiveness. We asked participants immediately after they

TABLE 3. Utility of the 20 Tests in the MATRICS Beta Battery as Repeated Measures: Performance Levels at Baseline and 4-Week Follow-Up

Domain and Tests	T_1 Mean	T_1 SD	T_2 Mean	T_2 SD	$T_2–T_1$ Difference Mean	$T_2–T_1$ Difference SD	Number of Scores at Floor/Ceiling	t	p	Effect Size (Cohen's d)
Speed of processing										
Category fluency test, animal naming	18.6	5.7	18.6	5.6	0.0	4.0	0/0	–0.04	0.96	0.00
Trail Making Test, Part A	42.7	15.6	39.5	14.4	–3.2	10.4	0/0	–4.09	<0.0001	0.22
Wechsler Adult Intelligence Scale, 3rd ed. (WAIS-III), digit symbol-coding subtest	52.3	13.4	55.5	15.0	3.2	7.8	0/0	5.24	<0.0001	0.22
Brief Assessment of Cognition in Schizophrenia (BACS), symbol coding subtest	41.2	12.1	41.6	11.3	0.4	6.4	0/0	0.74	0.46	0.03
Attention/vigilance										
3–7 Continuous Performance Test, shortened version	3.94	0.97	3.96	1.02	0.02	0.92	0/19	–0.65	0.52	0.04
Continuous Performance Test—Identical Pairs version	2.21	0.81	2.28	0.84	0.07	0.46	1/1	2.01	0.04	0.09
Working memory										
BACS, digit sequencing subtest	17.1	4.7	17.7	4.4	0.6	3.1	0/3	2.40	0.02	0.13
WAIS-III, letter-number sequencing subtest	8.7	2.9	8.9	2.6	0.2	1.9	0/0	1.32	0.19	0.07
Letter-Number Span test	11.6	4.1	12.3	3.7	0.7	2.4	0/0	3.28	0.001	0.16
Wechsler Memory Scale, 3rd ed., spatial span subtest	13.2	3.4	13.3	3.2	0.1	2.4	0/0	0.23	0.82	0.01
Spatial delayed response task	24.1	11.8	22.3	11.8	–1.8	8.2	0/0	–2.69	0.008	0.15
Verbal learning										
Neuropsychological Assessment Battery, daily living memory subtest	38.4	7.1	38.4	6.9	0.0	4.9	0/1	0.01	0.99	0.00
Hopkins Verbal Learning Test—Revised, immediate recall	21.3	5.4	21.8	5.7	0.5	4.3	0/0	1.51	0.13	0.09
Visual learning										
Neuropsychological Assessment Battery, shape learning subtest	14.3	4.6	14.0	4.7	–0.3	4.1	0/0	–1.00	0.32	0.07
Brief Visuospatial Memory Test—Revised	17.6	8.2	17.8	7.9	0.2	6.1	0/0	0.36	0.72	0.02
Reasoning and problem solving										
WAIS-III, block design subtest	29.4	13.3	32.5	13.7	3.1	7.0	1/0	5.62	<0.0001	0.22
BACS, Tower of London subtest	12.7	5.1	13.6	5.2	0.9	4.6	4/9	2.46	0.01	0.17
Neuropsychological Assessment Battery, mazes subtest	11.8	6.9	12.4	7.1	0.6	4.0	11/8	1.85	0.07	0.08
Social cognition										
Mayer-Salovey-Caruso Emotional Intelligence Test, perceiving emotions branch	94.7	16.6	93.9	18.3	–0.8	11.0	5/0	0.97	0.33	0.04
Mayer-Salovey-Caruso Emotional Intelligence Test, managing emotions branch	85.6	11.0	85.0	10.6	–0.6	8.0	0/0	0.95	0.34	0.06

took each test to point to a number on a 7-point Likert scale (indicated by unhappy to happy drawings of faces; 1=extremely unpleasant; 7=extremely pleasant) to indicate how unpleasant or pleasant they found the test.

Table 5 presents the practicality and tolerability results as well as the mean time it took to administer each test. Despite some variability, most tests were considered to be both practical and tolerable. This result likely reflects the efforts to take these factors into consideration in the earlier stages of test selection.

Selection of the Final Battery

The MATRICS Neurocognition Committee used the data in Tables 2–5 to select the tests that make up the final MCCB. Two site principal investigators of the Psychometric and Standardization Study who were not already part of the MATRICS Neurocognition Committee (Drs. Baade and Seidman) were added to the decision-making group to maximize input from experts in neurocognitive measurement. After a discussion of results in a given cognitive domain, committee members independently ranked each candidate test through an e-mail ballot. Members who had a conflict of interest with any test within a domain recused themselves from the vote on that domain. The 10 tests in the final battery are presented in Table 6 in the recommended order of administration. Based on the time the individual tests took to administer during the MATRICS Psychometric and Standardization Study, total testing time (without considering rest breaks) is estimated to be about 65 minutes. Training to administer these 10 tests should take no more than 1 day, including didactic in-

TABLE 4. Relationship of the 20 Tests in the MATRICS Beta Battery to Functional Outcome

| | Correlations[a] | | | | | | | |
| | Global | | Work Functioning | | Social Functioning | | Independent Living | |
Domain and Tests	Combined Data	Median	Combined Data	Median	Combined Data	Median	Combined Data	Median
Speed of processing								
Category fluency test, animal naming	0.18	0.21	0.09	0.13	0.04	0.09	0.18	0.04
Trail Making Test, Part A	0.25	0.25	0.30	0.32	0.07	0.08	0.06	0.06
Wechsler Adult Intelligence Scale, 3rd ed. (WAIS-III), digit symbol-coding subtest	0.31	0.38	0.28	0.26	0.10	0.19	0.16	0.15
Brief Assessment of Cognition in Schizophrenia (BACS), symbol coding subtest	0.28	0.31	0.29	0.30	0.03	0.13	0.18	0.18
Attention/vigilance								
3–7 Continuous Performance Test, shortened version	0.15	0.22	0.15	0.23	−0.04	−0.08	0.15	0.16
Continuous Performance Test—Identical Pairs version	0.25	0.22	0.24	0.34	−0.02	−0.06	0.21	0.18
Working memory								
BACS, digit sequencing subtest	0.09	0.06	0.18	0.18	−0.06	−0.05	0.04	0.10
WAIS-III, letter-number sequencing subtest	0.11	0.10	0.22	0.20	−0.08	−0.06	0.06	0.00
Letter-Number Span test	0.15	0.06	0.25	0.28	−0.04	0.06	0.05	0.03
Wechsler Memory Scale, 3rd ed., spatial span subtest	0.09	0.16	0.14	0.24	−0.02	−0.06	0.04	−0.01
Spatial delayed response task	0.17	0.18	0.24	0.27	−0.02	−0.09	0.09	0.02
Verbal learning								
Neuropsychological Assessment Battery, daily living memory subtest	0.18	0.12	0.15	0.17	0.03	−0.05	0.13	0.12
Hopkins Verbal Learning Test—Revised, immediate recall	0.14	0.22	0.09	0.14	−0.06	0.03	0.20	0.25
Visual learning								
Neuropsychological Assessment Battery, shape learning subtest	0.20	0.16	0.17	0.09	0.00	0.01	0.18	0.25
Brief Visuospatial Memory Test—Revised	0.20	0.22	0.17	0.26	0.11	0.10	0.07	0.05
Reasoning and problem solving								
WAIS-III, block design subtest	0.10	0.08	0.19	0.18	0.01	0.04	−0.02	−0.18
BACS, Tower of London subtest	0.25	0.27	0.20	0.20	0.11	0.15	0.13	0.22
Neuropsychological Assessment Battery, mazes subtest	0.15	0.18	0.21	0.17	0.12	0.18	−0.07	−0.05
Social cognition								
Mayer-Salovey-Caruso Emotional Intelligence Test, perceiving emotions branch	0.02	0.00	0.04	0.05	−0.11	−0.12	0.10	0.10
Mayer-Salovey-Caruso Emotional Intelligence Test, managing emotions branch	0.26	0.40	0.10	0.17	0.07	0.11	0.29	0.20

[a] Pearson's r; values are listed for the data combined across sites and for the median r of the five sites.

struction and hands-on practice. Below, we briefly summarize the reasons that these 10 tests were selected.

Speed of processing. The committee had planned to include two types of measures in this category: a verbal fluency measure and a graphomotor speed measure. The measures were psychometrically comparable in most respects. The Brief Assessment of Cognition in Schizophrenia symbol coding subtest was selected because it showed a smaller practice effect than the WAIS-III digit symbol coding subtest. Given the brief administration time and high tolerability of these measures, the committee decided to include an additional graphomotor measure with a different format (the Trail Making Test, Part A), for a total of three tests (the Trail Making Test, Part A; the Brief Assessment of Cognition in Schizophrenia symbol coding subtest; and the category fluency test).

Attention/vigilance. The Continuous Performance Test—Identical Pairs Version was selected for its high test-retest reliability and the absence of a ceiling effect.

Working memory. The spatial span subtest of the Wechsler Memory Scale, 3rd ed., was selected for nonverbal working memory because of its practicality, its brief administration time, and the absence of a practice effect. For verbal working memory, the Letter-Number Span test was selected because of its high reliability and its somewhat stronger relationship to global functional status.

Verbal learning. The committee considered the verbal learning tests to be psychometrically comparable. The Hopkins Verbal Learning Test—Revised was selected because of the availability of six forms, which may be helpful for clinical trials with several testing occasions.

TABLE 5. Practicality and Tolerability of the 20 Tests in the MATRICS Beta Battery

Domain and Tests	Practicality[a]					Tolerability			
	Global		Setup	Adminis-tration	Scoring	Participants' Ratings[b]		Administration Time	
	Mean	SD	Mean	Mean	Mean	Mean	SD	Mean	SD
Speed of processing									
Category fluency test, animal naming	6.4	1.0	7.0	7.0	6.7	5.4	1.1	2.0	0.4
Trail Making Test, Part A	6.9	0.3	7.0	7.0	6.9	5.6	1.2	2.1	0.7
Wechsler Adult Intelligence Scale, 3rd ed. (WAIS-III), digit symbol-coding subtest	6.7	0.7	6.8	6.9	6.3	5.2	1.2	3.4	1.1
Brief Assessment of Cognition in Schizophrenia (BACS), symbol coding subtest	6.7	0.7	6.9	6.9	5.9	5.5	1.1	3.0	0.5
Attention/vigilance									
3–7 Continuous Performance Test, shortened version	6.0	1.0	5.2	5.3	6.0	4.9	1.4	14.3	2.3
Continuous Performance Test—Identical Pairs version	5.7	1.1	5.2	4.1	5.7	4.2	1.6	13.4	2.1
Working memory									
BACS, digit sequencing subtest	6.8	0.4	6.9	6.7	6.9	4.4	1.3	6.2	1.7
WAIS-III, letter-number sequencing subtest	6.2	1.4	7.0	6.0	6.9	3.9	1.4	4.8	1.6
Letter-Number Span test	6.2	1.4	7.0	6.0	6.9	3.7	1.4	5.9	1.8
Wechsler Memory Scale, 3rd ed., spatial span subtest	5.3	1.2	6.6	5.2	6.1	4.8	1.4	5.1	1.7
Spatial delayed response task	2.9	0.8	3.8	2.1	3.6	4.8	1.4	16.4	3.4
Verbal learning									
Neuropsychological Assessment Battery, daily living memory subtest	6.2	1.3	6.2	6.4	6.4	4.6	1.4	4.6	1.0
Hopkins Verbal Learning Test—Revised, immediate recall	6.9	0.3	6.9	7.0	7.0	4.4	1.3	4.1	1.4
Visual learning									
Neuropsychological Assessment Battery, shape learning subtest	6.0	1.3	6.2	5.8	6.7	4.2	1.4	9.2	1.7
Brief Visuospatial Memory Test—Revised	5.4	0.9	6.4	5.9	3.0	4.4	1.5	4.7	1.4
Reasoning and problem solving									
WAIS-III, block design subtest	5.7	1.2	6.1	5.7	6.4	4.6	1.7	11.7	3.2
BACS, Tower of London subtest	5.6	1.2	6.1	4.2	6.9	5.0	1.4	6.9	1.7
Neuropsychological Assessment Battery, mazes subtest	6.2	0.8	6.3	5.6	6.8	4.6	1.6	11.2	3.2
Social cognition									
Mayer-Salovey-Caruso Emotional Intelligence Test, perceiving emotions branch	5.4	2.0	6.2	4.9	6.8	4.9	1.2	7.4	3.1
Mayer-Salovey-Caruso Emotional Intelligence Test, managing emotions branch	4.9	1.8	6.2	4.1	6.8	5.2	1.2	12.0	3.1

[a] Rated on a 7-point Likert scale, with 1=lowest and 7=highest.
[b] Rated on a 7-point Likert scale, with 1=extremely unpleasant and 7=extremely pleasant.

Visual learning. The Brief Visuospatial Memory Test—Revised was selected because it had higher test-retest reliability, a brief administration time, and the availability of six forms.

Reasoning and problem solving. The Neuropsychological Assessment Battery mazes subtest was selected for its high test-retest reliability, small practice effect, and high practicality ratings.

Social cognition. The managing emotions component of the Mayer-Salovey-Caruso Emotional Intelligence Test was selected for its relatively stronger relationship to functional status.

With selection of the final battery, it became possible to calculate test-retest reliabilities for cognitive domain scores that involve multiple tests and for an overall composite score for all 10 tests. Test scores were transformed to z-scores using the schizophrenia sample and averaged to examine the reliability of the composite scores. Four-week test-retest intraclass correlation coefficients were 0.71 for the speed of processing domain score, 0.85 for the working memory domain score, and 0.90 for the overall composite score.

Discussion

Discussions among representatives of the FDA, NIMH, and MATRICS indicated that the absence of a consensus cognitive battery had been an overriding obstacle to regulatory approval of any drug as a treatment for the core cognitive deficits of schizophrenia. The steps and data presented here moved the process from an initial nomination list of more than 90 cognitive tests to the selection of the final 10 tests. The process involved the participation of a large number of scientists in diverse fields to achieve a consensus of experts. To ensure a fair and effective process for selecting the best brief cognitive measures for clinical trials, the methods included both a structured consensus process to evaluate existing data and substantial new data collection.

One clear point to emerge from the MATRICS consensus meetings was the importance of reliable and valid assessment of cognitive functioning at the level of key cognitive domains (1, 4). Although some interventions may improve cognitive functioning generally, evidence from cognitive neuroscience and neuropharmacology suggests that many psychopharmacological agents may differentially target a subset of cognitive domains (35, 36). A battery assessing cognitive performance at the domain level requires more administration time than one that seeks to measure only global cognitive change. One challenge was to select a battery that adequately assessed cognitive domains and was still practical for large-scale multisite trials. The beta battery was twice as long as the final battery yet was well tolerated by participants, with 95% returning as requested for retesting. Furthermore, no participant who started a battery failed to complete it. The final battery offers reliable and valid measures in each of the seven cognitive domains and should be even easier to complete. If a clinical trial involves an intervention hypothesized to differentially affect specific cognitive domains, our experience with the beta battery suggests that the MCCB could be supplemented with additional tests to assess the targeted domains without substantial data loss through attrition.

The psychometric study sites varied substantially in the level and variance of their participants' functional status. In contrast, cognitive performance showed few significant site effects. Local factors may influence functional status in ways that are not attributable to cognitive abilities. Thus, cognitive abilities establish a potential for everyday functional level, while environmental factors (e.g., the local job market, employment placement aid, and housing availability) may influence whether variations in cognitive abilities manifest themselves in differing community functioning. In this instance, the sites varied in the extent to which treatment programs and local opportunities encouraged independent living and return to work, and this variation contributed to differences across sites in the magnitude of correlations between individual measures and functional outcome. To further examine variability across sites, we considered the correlation for each site be-

TABLE 6. The MATRICS Consensus Cognitive Battery

Test[a]	Domain
Trail Making Test, Part A	Speed of processing
Brief Assessment of Cognition in Schizophrenia, symbol coding subtest	Speed of processing
Hopkins Verbal Learning Test—Revised, immediate recall (three learning trials only)	Verbal learning
Wechsler Memory Scale, 3rd ed., spatial span subtest	Working memory (nonverbal)
Letter-Number Span test	Working memory (verbal)
Neuropsychological Assessment Battery, mazes subtest	Reasoning and problem solving
Brief Visuospatial Memory Test—Revised	Visual learning
Category fluency test, animal naming	Speed of processing
Mayer-Salovey-Caruso Emotional Intelligence Test, managing emotions branch	Social cognition
Continuous Performance Test, Identical Pairs version	Attention/vigilance

[a] Tests are listed in recommended order of administration.

tween the overall composite cognitive score and global functional status. Three sites showed clear relationships (with r values of 0.36, 0.38, and 0.44) and two did not (r values of 0.03 and 0.11). Calculating the correlations across sites probably led to a low estimate of the true magnitude of this correlation in people with schizophrenia. Nevertheless, the correlations allowed a reasonable direct comparison of the tests.

Another factor in the magnitude of correlations observed between cognitive performance and functional outcome may have been the restriction of community functioning measures to self-report data, which may include biases. Use of informants to broaden the information base or use of direct measures of functional abilities in the clinic may yield stronger relationships to cognitive performance (8).

The final report of the MATRICS initiative identified the components of the MCCB and recommended its use as the standard cognitive performance battery for clinical trials of potential cognition-enhancing interventions. This recommendation was unanimously endorsed by the NIMH Mental Health Advisory Council in April 2005, and it was accepted by the FDA's Division of Neuropharmacological Drug Products. To facilitate ease of use and distribution, the tests in the MCCB were placed into a convenient kit form by a nonprofit company, MATRICS Assessment, Inc. (37). The MCCB is distributed by Psychological Assessment Resources, Inc., Multi-Health Systems, Inc., and Harcourt Assessment, Inc.

The primary use of the MCCB is expected to be in clinical trials of potential cognition-enhancing drugs for schizophrenia and related disorders. Use of the MCCB in trials of cognitive remediation would facilitate comparison of results across studies. Another helpful application would be as a reference battery in basic studies of cognitive processes in schizophrenia, as it may aid evaluation of sample variations across studies.

The last step in the development of the MCCB was a five-site community standardization study to establish co-

norms for the tests, examine demographic correlates of performance, and allow demographically corrected standardized scores to be generated. This stage is described in the next article (7).

References

1. Green MF, Nuechterlein KH, Gold JM, Barch DM, Cohen J, Essock S, Fenton WS, Frese F, Goldberg TE, Heaton RK, Keefe RSE, Kern RS, Kraemer H, Stover E, Weinberger DR, Zalcman S, Marder SR: Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. Biol Psychiatry 2004; 56:301–307

2. Marder SR, Fenton WS: Measurement and treatment research to improve cognition in schizophrenia: NIMH MATRICS initiative to support the development of agents for improving cognition in schizophrenia. Schizophr Res 2004; 72:5–10

3. Kern RS, Green MF, Nuechterlein KH, Deng BH: NIMH-MATRICS survey on assessment of neurocognition in schizophrenia. Schizophr Res 2004; 72:11–19

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

5. Sergi MJ, Rassovsky Y, Nuechterlein KH, Green MF: Social perception as a mediator of the influence of early visual processing on functional status in schizophrenia. Am J Psychiatry 2006; 163:448–454

6. Brekke J, Kay DD, Lee KS, Green MF: Biosocial pathways to functional outcome in schizophrenia. Schizophr Res 2005; 80: 213–225

7. Kern RS, Nuechterlein KH, Green MF, Baade LE, Fenton WS, Gold JM, Keefe RSE, Mesholam-Gately R, Mintz J, Seidman LJ, Stover E, Marder SR: The MATRICS Consensus Cognitive Battery, part 2: co-norming and standardization. Am J Psychiatry (published online January 2, 2008; doi:10.1176/appi.ajp.2007.07010043)

8. Green MF, Nuechterlein KH, Kern RS, Baade LE, Fenton WS, Gold JM, Keefe RSE, Mesholam-Gately R, Seidman LJ, Stover E, Marder SR: Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS Psychometric and Standardization Study. Am J Psychiatry (published online January 2, 2008; doi:10.1176/appi.ajp.2007.07010089)

9. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDonnell J, Vader JP, Kahan JP: The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, Calif, RAND, 2001

10. Young AS, Forquer SL, Tran A, Starzynski M, Shatkin J: Identifying clinical competencies that support rehabilitation and empowerment in individuals with severe mental illness. J Behav Health Serv Res 2000; 27:321–333

11. Spreen O, Strauss E: A Compendium of Neuropsychological Tests. New York, Oxford University Press, 1991

12. Army Individual Test Battery: Manual of Directions and Scoring. Washington, DC, Adjutant General's Office, War Department, 1944

13. Wechsler D: Wechsler Adult Intelligence Scale—3rd ed (WAIS-III): Administration and Scoring Manual. San Antonio, Tex, Psychological Corp, 1997

14. Keefe RSE: Brief Assessment of Cognition in Schizophrenia (BACS) Manual—A: Version 2.1. Durham, NC, Duke University Medical Center, 1999

15. Nuechterlein KH, Edell WS, Norris M, Dawson ME: Attentional vulnerability indicators, thought disorder, and negative symptoms. Schizophr Bull 1986; 12:408–426

16. Cornblatt BA, Risch NJ, Faris G, Friedman D, Erlenmeyer-Kimling L: The Continuous Performance Test, Identical Pairs version (CPT-IP), I: new findings about sustained attention in normal families. Psychiatry Res 1988; 26:223–238

17. Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR: Auditory working memory and Wisconsin Card Sorting Test

performance in schizophrenia. Arch Gen Psychiatry 1997; 54: 159–165

18. Wechsler D: The Wechsler Memory Scale, 3rd ed. San Antonio, Tex, Psychological Corp (Harcourt), 1997

19. Hershey T, Craft S, Glauser TA, Hale S: Short-term and long-term memory in early temporal lobe dysfunction. Neuropsychology 1998; 12:52–64

20. White T, Stern RA: Neuropsychological Assessment Battery: Psychometric and Technical Manual. Lutz, Fla, Psychological Assessment Resources, Inc, 2003

21. Brandt J, Benedict RHB: The Hopkins Verbal Learning Test—Revised: Professional Manual. Odessa, Fla, Psychological Assessment Resources, Inc, 2001

22. Benedict RHB: Brief Visuospatial Memory Test—Revised: Professional Manual. Odessa, Fla, Psychological Assessment Resources, Inc, 1997

23. Mayer JD, Salovey P, Caruso DR: Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) User's Manual. Toronto, MHS Publishers, 2002

24. First MB, Spitzer RL, Gibbon M, Williams JBW: Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Edition (SCID-P), version 2. New York, New York State Psychiatric Institute, Biometrics Research, 1997

25. Ventura J, Lukoff D, Nuechterlein KH, Liberman RP, Green MF, Shaner A: Brief Psychiatric Rating Scale (BPRS), expanded version (4.0): scales, anchor points, and administration manual. Int J Methods Psychiatr Res 1993; 3:227–243

26. Overall JE, Gorham DR: The Brief Psychiatric Rating Scale. Psychol Rep 1962; 10:799–812

27. Birchwood M, Smith J, Cochran R, Wetton S, Copestake S: The social functioning scale: the development and validation of a new scale of social adjustment for use in family intervention programs with schizophrenic patients. Br J Psychiatry 1990; 157:853–859

28. Weissman M, Paykel E: The Depressed Woman: A Study of Social Relationships. Chicago, University of Chicago Press, 1974

29. Green MF, Kern RS, Heaton RK: Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. Schizophr Res 2004; 72:41–51

30. McKibbin CL, Brekke JS, Sires D, Jeste DV, Patterson TL: Direct assessment of functional abilities: relevance to persons with schizophrenia. Schizophr Res 2004; 72:53–67

31. Brekke JS, Long JD: Community-based psychosocial rehabilitation and prospective change in functional, clinical, and subjective experience variables in schizophrenia. Schizophr Bull 2000; 26:667–680

32. Brekke JS, Long JD, Nesbitt N, Sobel E: The impact of service characteristics from community support programs for persons with schizophrenia: a growth curve analysis. J Consult Clin Psychol 1997; 65:464–475

33. Kee KS, Green MF, Mintz J, Brekke JS: Is emotional processing a predictor of functional outcome in schizophrenia? Schizophr Bull 2003; 29:487–497

34. Gold JM, Goldberg RW, McNary SW, Dixon LB, Lehman AF: Cognitive correlates of job tenure among patients with severe mental illness. Am J Psychiatry 2002; 159:1395–1402

35. Geyer MA, Tamminga CA: Measurement and treatment research to improve cognition in schizophrenia: neuropharmacological aspects. Psychopharmacology (Berl) 2004; 174:1–2

36. Gazzaniga MS (ed): The Cognitive Neurosciences III, 3rd ed. Cambridge, Mass, MIT Press, 2004

37. Nuechterlein KH, Green MF: MATRICS Consensus Cognitive Battery. Los Angeles, MATRICS Assessment, Inc, 2006