Contents:

A. From Test Nominations to a Beta Battery

Methods and Results

Initial Evaluation

The initial step in moving from over 90 nominated cognitive tests to six or fewer candidate tests for each cognitive domain was accomplished through a series of conference calls among members of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Neurocognition Committee. The Neurocognition Committee was co-chaired by Drs. Nuechterlein and Green and included representatives from relevant fields in academia (Drs. Barch, Cohen, Essock, Gold, Heaton, Keefe, and Kraemer), NIMH (Drs. Fenton, Goldberg, Stover, Weinberger, and Zalcman), and consumer advocacy (Dr. Frese). Given the extensive experience of its members with cognitive assessment in schizophrenia, the Neurocognition Committee was able to do an initial evaluation of the extent to which each nominated test met the criteria developed at the first MATRICS consensus conference. Known reliability and validity of the tests in use with individuals with schizophrenia played a key role, as did feasibility for clinical trials. Because the initial survey established that the entire battery would optimally not require more than 90 minutes, individual tests were sought that could be completed with high reliability and validity in less than 15 minutes each. This initial review resulted in a list of 36 candidate tests across the seven cognitive domains.

Preparation of Database on Candidate Tests

The next steps in test selection used procedures based on the RAND/UCLA Appropriateness Method to increase consensus regarding selection of cognitive tests. This method was initially designed to develop criteria for measuring the quality of medical and surgical procedures (1). It has since evolved into a more general method, and it has been used in a wide variety of instances to increase consensus when the published research literature cannot provide definitive answers (2). The RAND/UCLA method starts with a broad review of all relevant scientific evidence. It

then uses, iteratively, methods that help increase agreement among members of an expert panel. Each panel member represents a key stakeholder group.

In this project, the first step in the RAND/UCLA method was to synthesize the published and unpublished relevant scientific evidence through a detailed literature review. Thus, for each cognitive domain, a summary of available information on each candidate test was compiled by the MATRICS staff at UCLA, including a test description and information on test-retest reliability, utility as a repeated measure, relationship to functional status, potential changeability in response to pharmacological agents, and practicality and tolerability. Published materials and consultation with test developers were used to compile this database. The resulting test database summaries can be accessed at www.matrics.ucla.edu (click on "Meetings & Presentations," then on "Conference 3").

Evaluation of Tests by Expert Panel

The second step of the RAND/UCLA consensus process involved an expert panel that evaluated the extent to which each candidate test met each criterion for test selection. The panel included experts on cognitive deficits in schizophrenia (Deanna Barch, Washington University, St. Louis; Eric Granholm, University of California, San Diego; Ann Kring, University of California, Berkeley; Barton Palmer, University of California, San Diego; Larry Seidman, Harvard Medical School), biostatistics and psychometrics (Helena Kraemer, Stanford University), clinical neuropsychology (Christopher Randolph, Chicago Neurological Institute), clinical trials methodology (Georges Gharabawi, then at Janssen Pharmaceutica; Donald Goff, Harvard Medical School), cognitive science (Edward E. Smith, then at University of Michigan), neuropharmacology (Alan Breier, Eli Lilly; Steven Johnson, Cortex Pharmaceuticals; Christopher Schmidt, Pfizer), and clinical psychiatry (Alexander Miller, University of Texas Health Sciences Center, San Antonio).

Using the database, each expert initially made an independent rating of each of the 36 tests on each of the five test selection criteria separately, using a 9-point scale (180 ratings). The anchor points ranged from poor to superb. Experts were provided with recommended considerations when making judgments about a criterion (see Conference 3 materials at www.matrics.ucla.edu). These preconference ratings were examined to identify instances of dispersion of ratings that indicate lack of consensus, using SD>1.5 and notable bimodality as guides. Twenty of the 180 ratings showed notable lack of consensus.

The expert panel met in Los Angeles in September 2003 and discussed in detail the ratings that showed notable lack of consensus. The database summaries for each of these 20 ratings were discussed in turn. These 20 ratings were then confidentially completed again. Dispersion of these ratings decreased, and medians for nine ratings changed. The medians of the resulting final ratings are presented in Table 1 in the article, grouped by cognitive domain.

The expert panel also recommended that one criterion, potential changeability in response to pharmacological agents, not be used in selecting the final battery. Data on sensitivity to pharmacological agents were often sparse and were usually not based on agents specifically developed to produce cognitive change. The panel did not believe that this criterion could provide useful guidance about the sensitivity of cognitive measures to drugs that would be the focus of clinical trials.

Selection of Beta Version of Battery

The results from the expert panel were reviewed by the Neurocognition Committee and used to select measures for inclusion in the beta version of the battery for the MATRICS Psychometric and Standardization Study (PASS). The goal was to select two to four measures from each of the seven neurocognitive domains, depending on test duration and content sampling considerations. For working memory, the substantial correlation of verbal and nonverbal working memory performance in schizophrenia, combined with the desirability of including a nonverbal measure to facilitate appropriate animal models, had led to a decision to have only one working memory domain but to include a verbal and a nonverbal working memory test in the final battery (3). Thus, enough working memory measures were included to allow the best candidates for a verbal and a nonverbal working memory test to be evaluated. As part of this process, the original Letter-Number Span test was added as a verbal working memory measure for comparison with the specific version (letter-number sequencing) used in the WAIS-III. Measures of speed of processing were very brief, so four were included. Three rather than two tests were included for reasoning and problem solving to broaden the content representation of that domain.

The Neurocognition Committee followed the ratings of the expert panel members in making the final selections unless there were special circumstances for ranking tests out of order. This occurred in one of the seven domains, verbal learning, in which a list learning test that had more alternate forms (the Hopkins Verbal Learning Test—Revised) was selected, even though it

had not been rated quite as highly as other list learning tests that had one alternate form. The committee believed that clinical trials benefit from the option to use more than one form in domains in which practice effects would be expected.

The resulting beta version of the MATRICS Consensus Cognitive Battery included 20 tests. Descriptions of these tests are available from their associated references, or on the MATRICS web site (see Conference 3 materials at www.matrics.ucla.edu). The tests were as follows:

Speed of processing: Category fluency test, animal naming (4); Trail Making Test, Part A (5); Wechsler Adult Intelligence Scale, 3rd ed. (WAIS-III), digit symbol-coding subtest (6); and Brief Assessment of Cognition in Schizophrenia (BACS), symbol coding subtest (7)

Attention/vigilance: Shortened version of the 3–7 Continuous Performance Test (8), and Continuous Performance Test—Identical Pairs version (9)

Working memory: verbal working memory: BACS, digit sequencing subtest (7); WAIS-III, letter-number sequencing subtest (6); and Letter-Number Span test (10). *Nonverbal working memory*: Wechsler Memory Scale, 3rd ed. (WMS-III), spatial span subest (11); and spatial delayed response task (12)

Verbal learning: Neuropsychological Assessment Battery, daily living memory subtest (13); and Hopkins Verbal Learning Test—Revised, immediate recall subtest (14)

Visual learning: Neuropsychological Assessment Battery, shape learning subtest (13); and Brief Visuospatial Memory Test—Revised (15)

Reasoning and problem solving: WAIS-III, block design subtest (6); BACS, Tower of London subtest (7); and Neuropsychological Assessment Battery, mazes subtest (13)

Social cognition: Mayer-Salovey-Caruso Emotional Intelligence Test, perceiving emotions branch (16); and Mayer-Salovey-Caruso Emotional Intelligence Test, managing emotions branch (16)

B. From Beta Battery to Final Battery

Method

To move from the beta version to the final version of the battery, a five-site study, MATRICS PASS, was conducted to directly compare the tests on their psychometric properties (including test-retest reliability, practice effects, and relationships to functional status) and practicality/tolerability. The tests were compared within each cognitive domain so that the best representative(s) of each domain would be selected for the final battery.

Participants

The sample was intended to represent individuals with schizophrenia who might be included in a typical psychopharmacology clinical trial of a cognition-enhancing agent. Each of the five sites contributed at least 30 participants with schizophrenia (Ns of 33 to 37 across sites) who were tested twice, 4 weeks apart. We accepted participants who were currently taking a variety of medications, but we excluded individuals who had a history or condition aside from schizophrenia that might influence performance on cognitive measures. Participants were selected according to the following criteria: diagnosis of schizophrenia or schizoaffective disorder, depressed type, based on diagnostic interview; no medication changes in previous month and none anticipated for the following month; stable outpatient or rehabilitation center status; age 18–65 years; no substance dependence in past 6 months; no substance abuse in past month; no clinically significant neurological disease or head injury as determined by medical history; ability to understand spoken English sufficiently to validly complete testing procedures; and ability to comprehend the consent form appropriately. We also excluded individuals who did not meet the substance use or dependence exclusion criteria but who had a clearly excessive amount of lifetime alcohol or drug consumption over a 10-year period or had been using alcohol or drugs heavily in the 3 days prior to testing (rating instrument from L.J. Seidman and W.S. Kremen, unpublished manuscript; lseidman@bidmc.harvard.edu).

Sites

Performance sites were selected that had extensive experience with conducting clinical trials with schizophrenia samples and also had local expertise in neuropsychological assessment. Final

site selection was based on discussions with the NIMH project officer for MATRICS (Dr. Fenton) and represented diverse geographic regions. The five sites and their principal investigators and co-principal investigators were 1) UCLA and the VA Greater Los Angeles Healthcare System (coordinating site) (Drs. Green and Nuechterlein); 2) Duke University (Dr. Keefe); 3) Maryland Psychiatric Research Center, University of Maryland (Dr. Gold); 4) Massachusetts Mental Health Center, Harvard Medical School (Drs. Seidman and Mesholam-Gately); and 5) University of Kansas (Dr. Baade).

Study design

Potential participants received a complete description of the study and provided informed consent. Human subject procedures were approved by each site's institutional review board. In addition, because UCLA was the coordinating site, the consent forms from all sites were reviewed by the UCLA institutional review board. After consenting, potential participants received a diagnostic interview with the Structured Clinical Interview for DSM-IV (17), administered by an experienced interviewer. If the individual met study entry criteria, he or she was scheduled for baseline assessments. Participants were asked to return 4 weeks after baseline assessment for a retest.

Assessments

In addition to the cognitive performance measures, data were collected on clinical symptoms to characterize the sample, and self-report measures of community functioning were collected to examine correlations with cognitive measures. Although evaluation of cognitive performance measures was the focus of this project, we included measures of functional capacity and interview-based measures of cognition in MATRICS PASS (18, 19) to evaluate their potential use as coprimary measures in clinical trials (20).

Cognitive performance measures

The 20 tests in the beta version of the battery were administered. The test administrators were research assistants, usually with a bachelor's degree, who had prior experience in administering standardized cognitive tests. A 2-day training session was held in Los Angeles to ensure that all tests were administered and scored in the standardized way. Training included didactic sessions and hands-on practice led by experienced psychologists. Regular conference calls among all test

administrators and the neuropsychological testing coordinator (Dr. Robert Kern) were used to prevent drift in administration procedures over time.

Five of the tests in the beta battery have alternate forms: the Hopkins Verbal Learning Test—Revised, the Brief Visuospatial Memory Test—Revised, and the three tests from the Neuropsychological Assessment Battery (daily living memory, shape learning, and mazes subtests). The inclusion of alternate forms in the psychometric evaluation has advantages as well as disadvantages. Alternate forms limit practice effects and generally help prevent performance from reaching ceiling. However, alternate forms are not perfectly equivalent, and this may introduce a source of variability (form-to-form variability) and a potential drop in test-retest reliability. In PASS, we decided to use alternate forms when available because we thought that if these tests were selected for the final battery, the alternate forms would likely be used in many clinical trials. For the Neuropsychological Assessment Battery subtests, the one available alternate form was used. The Hopkins Verbal Learning Test—Revised and the Brief Visuospatial Memory Test—Revised both have six alternate forms. Based on consultation with the test publisher and test developer, we selected form 4 of the Hopkins Verbal Learning Test—Revised and form 5 of the Brief Visuospatial Memory Test—Revised. Administration of the alternate forms was counterbalanced so that half the patients received form 1 first, and half received the alternate form.

Symptom rating scales

The Brief Psychiatric Rating Scale (BPRS) (21), a semistructured instrument, was used to rate the presence and severity of psychiatric symptoms (range, 1–7 for each item). The 24-item version of the BPRS (22) was administered at the baseline and follow-up assessments. Clinical interviewers were trained on this scale based on a modified training and QA program used in the VISN 22 Mental Illness Research, Education, and Clinical Center Treatment Unit. Each trainee attended a 2-day training session in Los Angeles, where they completed role-played interviews and rated six training videos. After returning to their home site, each trainee completed and videotaped two interviews. The trainer then co-rated these tapes and provided feedback to the trainee. On the co-rated interviews, trainees were required to obtain a minimum median intraclass correlation coefficient with gold standard ratings of 0.80 across items (23).

Community functioning

To assess community functioning, we used the Birchwood Social Functioning Scale (24), a measure that was recommended by the MATRICS Outcomes Committee (Alan Bellack, Chair). This scale includes self-report questions about social engagement/withdrawal, interpersonal behavior, prosocial activities, recreation, independence-competence, independence-performance, and employment/occupation. To create more uniform rating methods across sites, anchor points were added to the scale and it was administered in an interview format. Because the Social Functioning Scale has relatively few items about employment and school functioning, it was supplemented with work and school items from the Social Adjustment Scale (25). While prior studies of the relationships between cognitive performance and functional outcome have used a range of measures of functional skills and functioning levels (26), for this evaluation self-report measures of community functioning were separated from laboratory-based measures of functional capacity (19), as the latter were being considered as possible coprimary measures for clinical trials (20).

Ratings for all items involved the previous 3-month period. Training in community functioning ratings occurred during the same 2-day initial meeting as other measures and included detailed discussion of the anchor points. To ensure that the ratings of community functioning were not influenced by knowledge of cognitive performance, different individuals administered the cognitive performance and community functioning measures.

C. Supplementary Tables

Supplementary Table 1. Factor Loadings for Functional Status Variables[a]

Items	Factor 1: Social Functioning	Factor 2: Independent Living	Factor 3: Work Functioning
Current living situation	–0.23	**0.70**	0.10
Social engagement/withdrawal	**0.63**	–0.12	0.18
Interpersonal communication	**0.58**	0.35	–0.04
Independence performance	0.48	**0.64**	0.25
Independence competence	0.15	**0.73**	0.09
Recreation	**0.65**	0.13	0.01
Prosocial activities	**0.71**	–0.06	0.17
Employment index	0.19	0.20	**0.83**
Hours at school or work	0.06	0.08	**0.90**

[a] Bold font identifies the primary contributors to each factor.

Supplementary Table 2. Test-Retest Reliability of the 20 Tests in the Beta Version of the MATRICS Consensus Cognitive Battery (Pearson's r and Intraclass Correlation Coefficient [ICC])

			Test-Retest Reliability	
Domain	Test	Test Scores Used	r	ICC
Speed of processing	Category fluency test, animal naming	Total number of animals named in 60 seconds	0.74	0.74
	Trail Making Test, Part A	Time to completion	0.77	0.75
	Wechsler Adult Intelligence Scale–III (WAIS-III), digit symbol-coding subtest	Total number correct	0.85	0.83
	Brief Assessment of Cognition in Schizophrenia (BACS), symbol coding subtest	Total number correct	0.85	0.85
Attention/ vigilance	3–7 Continuous Performance Test, shortened version	Overall d′	0.60	0.60
	Continuous Performance Test—Identical Pairs version	Mean d′ value across 2-, 3-, and 4-digit conditions	0.84	0.84
Working memory	BACS, digit sequencing subtest	Number of correct responses	0.77	0.75
	WAIS-III, letter-number sequencing subtest	Number of correct trials	0.76	0.75
	Letter-Number Span test	Number of correct trials	0.81	0.78
	Wechsler Memory Scale, 3rd ed., spatial span subtest	Sum of raw scores on forward and backward conditions	0.74	0.74
	Spatial delayed response task	Distance from presented dot to remembered dot	0.75	0.73
Verbal learning	Neuropsychological Assessment Battery, daily living memory subtest	Total correct free recall across three trials	0.75	0.74
	Hopkins Verbal Learning Test—Revised, immediate recall	Total number of words recalled correctly over three learning trials	0.69	0.68
Visual learning	Neuropsychological Assessment Battery, shape learning subtest	Total learning score over three trials	0.61	0.61
	Brief Visuospatial Memory Test—Revised	Total recall score over three learning trials	0.71	0.71
Reasoning and problem solving	WAIS-III, block design subtest	Total raw score	0.87	0.84
	BACS, Tower of London subtest	Number of correct trials	0.59	0.58
	Neuropsychological Assessment Battery, mazes subtest	Total raw score	0.83	0.83
Social cognition	Mayer-Salovey-Caruso Emotional Intelligence Test, perceiving emotions branch	Branch score using general consensus scoring	0.80	0.80
	Mayer-Salovey-Caruso Emotional Intelligence Test, managing emotions branch	Branch score using general consensus scoring	0.73	0.73

References

1. 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
2. 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
3. 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
4. Spreen O, Strauss E: A Compendium of Neuropsychological Tests. New York, Oxford University Press, 1991
5. Army Individual Test Battery: Manual of Directions and Scoring. Washington, DC, War Department, Adjutant General's Office, 1944
6. Wechsler D: Wechsler Adult Intelligence Scale, 3rd ed (WAIS-III): Administration and Scoring Manual. San Antonio, Tex, Psychological Corp, 1997
7. Keefe RSE: Brief Assessment of Cognition in Schizophrenia (BACS) Manual—A, version 2.1. Durham, NC, Duke University Medical Center, 1999
8. Nuechterlein KH, Edell WS, Norris M, Dawson ME: Attentional vulnerability indicators, thought disorder, and negative symptoms. Schizophr Bull 1986; 12:408–426
9. 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. Psychiatr Res 1988; 26:223–238
10. 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
11. Wechsler D: The Wechsler Memory Scale, 3rd ed. San Antonio, Tex, Psychological Corp (Harcourt), 1997
12. Hershey T, Craft S, Glauser TA, Hale S: Short-term and long-term memory in early temporal lobe dysfunction. Neuropsychology 1998; 12:52–64
13. White T, Stern RA: Neuropsychological Assessment Battery: Psychometric and Technical Manual. Lutz, Fla, Psychological Assessment Resources, Inc, 2003

14. Brandt J, Benedict RHB: The Hopkins Verbal Learning Test—Revised: Professional Manual. Odessa, Fla, Psychological Assessment Resources, Inc, 2001
15. Benedict RHB: Brief Visuospatial Memory Test—Revised: Professional Manual. Odessa, Fla, Psychological Assessment Resources, Inc, 1997
16. Mayer JD, Salovey P, Caruso DR: Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) User's Manual. Toronto, MHS Publishers, 2002
17. 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
18. Green MF, Kern RS, Heaton RK: Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. Schizophr Res 2004; 72:41–51
19. 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
20. Green MF, Nuechterlein KH, Kern RS, Baade LE, Fenton WS, Gold JM, Keefe RSE, Mecholam-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)
21. Overall JE, Gorham DR: The Brief Psychiatric Rating Scale. Psychol Rep 1962; 10:799–812
22. 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
23. 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
24. 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
25. Weissman M, Paykel E: The Depressed Woman: A Study of Social Relationships. Chicago, University of Chicago Press, 1974
26. 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