

The MATRICS Consensus Cognitive Battery, Part 2: Co-Norming and Standardization

Robert S. Kern, Ph.D.

Keith H. Nuechterlein, Ph.D.

Michael F. Green, Ph.D.

Lyle E. Baade, Ph.D.

Wayne S. Fenton, M.D.

James M. Gold, Ph.D.

Richard S.E. Keefe, Ph.D.

Raquelle Mesholam-Gately, Ph.D.

Jim Mintz, Ph.D.

Larry J. Seidman, Ph.D.

Ellen Stover, Ph.D.

Stephen R. Marder, M.D.

Objective: The consensus cognitive battery developed by the National Institute of Mental Health's (NIMH's) Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative includes 10 independently developed tests that are recommended as the standard battery for clinical trials of cognition-enhancing interventions for schizophrenia. To facilitate interpretation of results from the MATRICS Consensus Cognitive Battery using a common scaling across tests, normative data were obtained from a single representative U.S. community sample with the battery administered as a unit.

Method: The MATRICS Consensus Cognitive Battery was administered to 300 individuals from the general community at five sites in differing geographic regions. For each site, recruitment was stratified by age, gender, and education. A scientific survey sampling method was used to help avoid sampling bias. The battery was administered in a standard order to each

participant in a single session lasting approximately 60 minutes. Descriptive data were generated, and age, gender, and education effects on performance were examined.

Results: Prominent age and education effects were observed across tests. The results for gender differed by measure, suggesting the need for age and gender corrections in clinical trials. The MATRICS Consensus Cognitive Battery components were co-normed, with allowance for demographic corrections.

Conclusions: Co-norming a battery such as the MATRICS Consensus Cognitive Battery, comprising tests from independent test developers each with their own set of norms, facilitates valid interpretation of test scores and communication of findings across studies. These normative data will aid in estimating the magnitude of change during clinical trials of cognition-enhancing agents and make it possible to derive more directly interpretable composite scores.

(*Am J Psychiatry* 2008; 165:214–220)

A key deliverable of the National Institute of Mental Health's (NIMH's) Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) contract was the development of a consensus battery to assess cognition in clinical trials of schizophrenia. This task was accomplished through a series of consensus conferences sponsored by MATRICS (1–3). The preceding article in this issue (4) describes the steps leading to the final consensus battery. Phase 2 of the MATRICS Psychometric and Standardization Study was a standardization study to obtain normative data on the MATRICS Consensus Cognitive Battery (MCCB) and examine age, gender, and education effects on performance.

The MCCB is a hybrid battery comprising multiple independently owned and published tests, which is the type of battery most commonly used in clinical trials research and clinical practice (5, 6). The importance of the availability of normative data for hybrid batteries has been highlighted recently (7, 8). It is argued that the normative data for such batteries must come from a single representative

sample to which the tests are administered together as a unit ("co-norming"). A problem with batteries made up of separately developed tests is that the normative reference samples for the individual tests vary widely in size and composition and hence are not directly comparable (9). Estimates of cognitive performance levels can vary across tests because of differences in the normative samples or test administration procedures, as opposed to actual differences in performance. For some tests, multiple sets of norms are available that come from published reports of independent studies with different sample compositions, leaving the selection of which norms to use up to the discretion of the investigator or clinician.

Another concern is that interpretation of findings in batteries that use multiple independently developed tests may be suspect because of a lack of normative information about the behavior of the battery as a whole. The normative information about the base rates of test score differences, test score variance, and the covariance among tests that is necessary for valid interpretation is generally not available

This article is featured in this month's AJP [Audio](#) and discussed in an editorial by Drs. Harvey and Cornblatt on p. 163.

for such batteries, a situation that can lead to faulty conclusions about the clinical meaning of test results.

For these reasons, the MATRICS Neurocognition Committee concluded that dependable and valid interpretation of test results from the MCCB would require a standardization and co-norming process. Surveyed experts viewed the presence of appropriate norms as an essential characteristic for a cognitive battery to be used in clinical trials (10). Such norms add clinical meaningfulness to study results; make it easier to communicate findings to a broader audience; facilitate communication of test findings across clinical trials because studies will use the same normative data to derive test scores; allow construction of a cognitive performance profile across the domains represented in the battery, as the domain scores can use a common metric; and facilitate the combination of test scores to yield a single index (an overall composite score) of cognitive functioning. The experts who participated in the MATRICS consensus meetings agreed that test norms should be representative of the larger population and thus should come from sufficiently large community samples that are representative of age, gender, educational, ethnicity, and regional distributions in the United States. In this article, we present the normative data from phase 2 of the MATRICS Psychometric and Standardization Study, which entailed the administration of the MCCB to a community standardization sample.

Method

Participants

Community participants were recruited at the same five sites as the patient sample in phase 1 of the MATRICS Psychometric and Standardization Study. The sites offered diversity in U.S. geographic region, urban/rural settings, and racial/ethnic background of participants. The sites represented the Northeast (Harvard Medical School, Boston), the East (University of Maryland, Baltimore), the South (Duke University, Durham, N.C.), the Midwest (University of Kansas, Wichita), and the West (University of California, Los Angeles, and the VA Greater Los Angeles Healthcare System). Each of the five sites contributed at least 60 participants. Of 312 participants, 12 who fell into categories that were overrepresented in a stratification by age, gender, and education were randomly dropped. The final sample included 300 participants, 100 from each of three age groups.

The sample was intended to approximate the gender, education, and racial/ethnic composition described in the 2000 report of the U.S. Census Bureau (11). At each site, recruitment was stratified by age, gender, and education. Because age effects are fairly large for cognitive tests, age was chosen as the primary stratification variable. For age, equal numbers of participants were drawn from three age groups: 20–39 years, 40–49 years, and 50–59 years. Because age-related changes in cognition are typically small for persons in their 20s and 30s (12), these two decades were treated as a single age group. For gender, approximately equal numbers of men and women were sought; the final sample included 47% men (53%, 46%, and 41% across the three age groups, respectively). For education, the sample was stratified according to three groups to approximate the U.S. distribution: less than a high school degree, at least a high school degree but less than a bachelor's degree, and a bachelor's degree or higher. The final sample

yielded 8% with less than a high school degree, 58% with a high school degree but less than a college degree, and 34% with at least a bachelor's degree. These proportions were consistent across age groups and similar to those reported for the 2000 U.S. census. Although the sample was not stratified for racial and ethnic group representation, racial and ethnic diversity was sought. The racial distribution of the final sample was 76% white, 18% African American, 2% Asian or Pacific Islander, and 4% other; 6% of participants were of Hispanic or Latino ethnicity. The basic demographic characteristics of the community sample are summarized in Table 1.

Recruitment Procedures

To help avoid the kind of sampling bias that can occur when study participants are recruited through flyers and published advertisements, the recruitment procedures adhered to a scientific survey sampling method. Zip codes around each site were selected on the basis of the desired diversity of demographic characteristics. Lists of randomly sampled residential telephone numbers in these zip code areas were purchased from Survey Sampling, Inc. (Fairfield, Conn.). Research staff members called these numbers and used a prepared script approved by each local institutional review board that explained the purpose of the study and asked whether respondents would consider participating. Based on the answers from responding households, staff members did preliminary telephone screening for demographic characteristics and general exclusion criteria. Persons who passed this initial screening and who were interested in participating were scheduled for an in-person interview to ensure that entry requirements were satisfied (see below). Persons meeting study entry criteria were then tested at the respective university or hospital research sites, generally immediately after the in-person interview.

Potential participants were excluded if they met any of the following criteria: history of diagnosis of schizophrenia or other psychotic disorder; clinically significant neurological disease or head injury with loss of consciousness lasting more than 1 hour, as determined by medical history; diagnosis of mental retardation or pervasive developmental disorder; currently taking an antipsychotic, antidepressant, antianxiety, or cognition-enhancing medication (e.g., methylphenidate) or, in the last 72 hours, narcotics for pain; history of heavy, sustained abuse of alcohol or drugs for a period of 10 or more years (13); any substance use for each of the last 3 days prior to testing; more than four alcoholic drinks per day for each of the last 3 days prior to testing; inability to understand spoken English sufficiently to comprehend testing instructions; and inability to comprehend the consent form appropriately.

Study Design

Participants provided written informed consent after the study procedures were fully explained. They were then formally screened using a structured interview by an experienced interviewer to determine whether they met the study's inclusion and exclusion criteria. The interview included questions about demographic characteristics and work/school status, history of head injury, neurological and psychiatric disorders, currently prescribed medications, and alcohol and substance abuse. The 10 tests of the MCCB (14–23) were administered to participants who met study criteria.

Three additional measures from the beta battery that were not part of the final battery were also administered: the Neuropsychological Assessment Battery, daily living memory subtest (19); the Brief Assessment of Cognition in Schizophrenia, Tower of London subtest (15); and the Neuropsychological Assessment Battery, shape learning subtest (19). These supplemental tests were included to gather additional normative data on three domains that each only had one representative test in the battery: verbal learn-

TABLE 1. Demographic Characteristics of a Community Sample (N=300), Stratified by Age, Gender, and Education, for the Standardization and Co-Norming of the MATRICS Consensus Cognitive Battery

Variable	Age Group					
	20–39 Years (N=100)		40–49 Years (N=100)		50–59 Years (N=100)	
	Mean	SD	Mean	SD	Mean	SD
Age (years)	28.4	5.7	45.3	2.8	54.4	2.6
Education (years)	14.5	2.5	14.5	2.4	14.2	3.0
	N	%	N	%	N	%
Male	53	53	46	46	41	41
White	82	82	74	74	72	72

ing; reasoning and problem solving; and visual learning. The decision to include these supplemental tests was based on a review of their psychometric data and the putative added value when administered in conjunction with the 10 MCCB tests. The MATRICS Neurocognition Committee did not recommend that any of the supplemental tests be used instead of the tests in the MCCB, only in addition to them.

The tests of the MCCB were administered in the same order to all participants (see supplementary Table 1 in the data supplement that accompanies the online edition of this article for the list of tests and their corresponding primary scores), and the supplemental tests were always administered after the completion of the MCCB. The final administration order was determined by two considerations: 1) starting the battery with less cognitively demanding tasks that were relatively straightforward and easy to understand, and 2) alternating verbal measures with nonverbal ones to alleviate processing burden and minimize interference among tests. Testing was conducted on one occasion, requiring approximately 1.5 hours (60 minutes for the MCCB plus 30 minutes for the supplemental measures). Participants were reimbursed for their time (including travel, in-person screening, and testing).

Data Analysis

The primary scores for each of the 10 tests in the MCCB were initially examined for normality of distribution. Distributions that were notably skewed were transformed. All raw or transformed raw scores were standardized to T scores based on the sample of 300 community participants. For cognitive domains that included more than one measure, the summary score for the domain was calculated by summing the T scores of the tests included in that domain and then standardizing the sum to a T score. An overall composite score for global cognition was calculated similarly: the seven domain T scores were summed, and the overall composite score was calculated by standardizing this sum to a T score based on the community sample. With the scores calculated in this way, all test scores, domain scores, and the overall composite score were standardized to the same measurement scale with a mean of 50 and an SD of 10. For composite scores calculated by simply taking the mean of a set of standardized T scores, although the SD of the individual scores is scaled to be 10, the SD of the mean of sets of such scores is not necessarily 10 and is generally lower (7). The magnitude of the SDs is affected by the number of measures included in the composite and their inter-correlations. In evaluating cognitive performance, interpretation is aided if the SDs are the same for individual and composite measures. The procedure used to compute composite scores in the MCCB yields equivalent SDs for all measures.

The T scores were analyzed to examine age, gender, and education effects. Differences in performance across the seven cognitive domains and the overall composite score were assessed using one-way analyses of variance (ANOVAs) or, in the case of gender effects, independent t tests. Follow-up contrasts were conducted for significant age and education effects, with a p value <0.05 considered statistically significant.

Results

Examination of the 10 raw score distributions for the MCCB revealed that three were notably skewed: the Trail Making Test, Part A; the Hopkins Verbal Learning Test—Revised; and the Neuropsychological Assessment Battery mazes subtest. Logarithmic transformations were performed on these scores. For the Trail Making Test, the log transformed score was reversed so that high values indicated good performance, thus making it directionally consistent with the other measures of speed of processing for the purpose of deriving a summary score for this domain. The nontransformed raw score means and SDs for the MCCB and supplemental tests are presented by age group in supplementary Table 2 in the online data supplement.

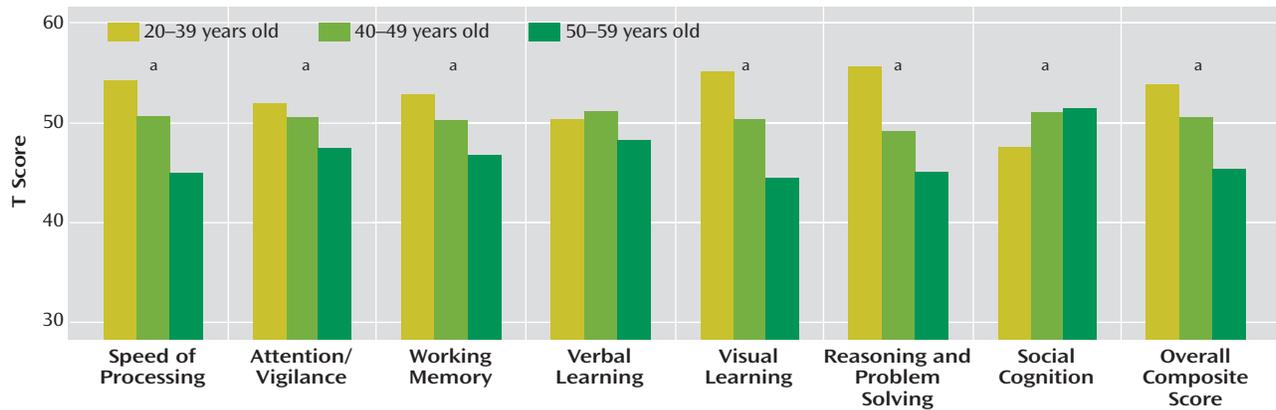
Age Effects

The one-way ANOVAs to assess age effects revealed a significant overall effect for six of the seven cognitive domains plus the overall composite score (see Figure 1; speed of processing: $F=25.45$, $df=2$, 296, $p<0.001$; attention/vigilance: $F=5.39$, $df=2$, 293, $p<0.005$; working memory: $F=10.00$, $df=2$, 295, $p<0.001$; visual learning: $F=34.12$, $df=2$, 296, $p<0.001$; reasoning and problem solving: $F=34.54$, $df=2$, 295, $p<0.001$; social cognition: $F=4.56$, $df=2$, 294, $p<0.001$; overall composite score: $F=19.25$, $df=2$, 287, $p<0.001$). The only cognitive domain that did not show a significant age effect was verbal learning, which revealed a nonsignificant trend in the expected direction. The pattern of age effects was similar across cognitive domains, with the exception of social cognition. In general, younger participants performed better than older ones (see the online data supplement for follow-up contrasts).

Gender Effects

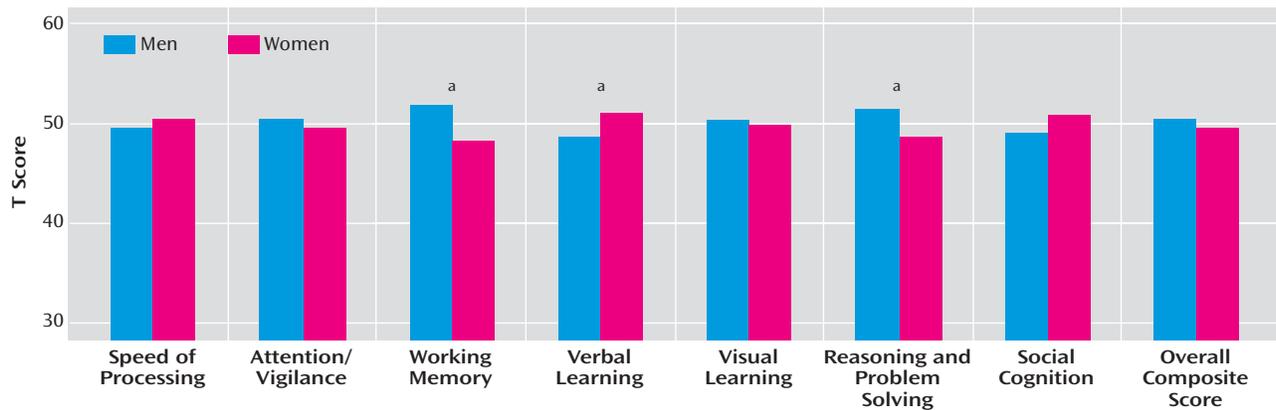
The analyses examining gender effects yielded a mixed pattern of results. Significant differences between men and women were observed in three of the seven cognitive domains (Figure 2). Men performed better than women on the measures for reasoning and problem solving ($t=2.45$, $df=296$, $p<0.02$) and for working memory ($t=3.09$, $df=296$, $p<0.01$). Women performed better than men on the verbal learning measure ($t=-2.05$, $df=298$, $p<0.05$). There was no statistically significant difference between men and women on the other four cognitive domains or the overall composite score.

FIGURE 1. Standardization Sample's Mean T Scores on the Cognitive Domains of the MATRICS Consensus Cognitive Battery, by Age Group



^a Significant group effect, $p < 0.05$.

FIGURE 2. Standardization Sample's Mean T Scores on the Cognitive Domains of the MATRICS Consensus Cognitive Battery, by Gender



^a Significant group effect, $p < 0.05$.

Education Effects

Education effects were highly consistent across all cognitive domains. The one-way ANOVAs revealed a significant overall effect on all seven cognitive domains plus the overall composite score (see Figure 3; speed of processing: $F=20.69$, $df=2$, 296 , $p < 0.001$; attention/vigilance: $F=31.17$, $df=2$, 293 , $p < 0.001$; working memory: $F=15.10$, $df=2$, 295 , $p < 0.001$; verbal learning: $F=14.65$, $df=2$, 297 , $p < 0.001$; visual learning: $F=13.92$, $df=2$, 296 , $p < 0.001$; reasoning and problem solving: $F=4.24$, $df=2$, 295 , $p < 0.02$; social cognition: $F=3.43$, $df=2$, 294 , $p < 0.04$; overall composite score: $F=22.31$, $df=2$, 287 , $p < 0.001$). The pattern of performance differences was highly similar across cognitive domains, with higher test performance corresponding with higher levels of education (see the online data supplement for follow-up contrasts).

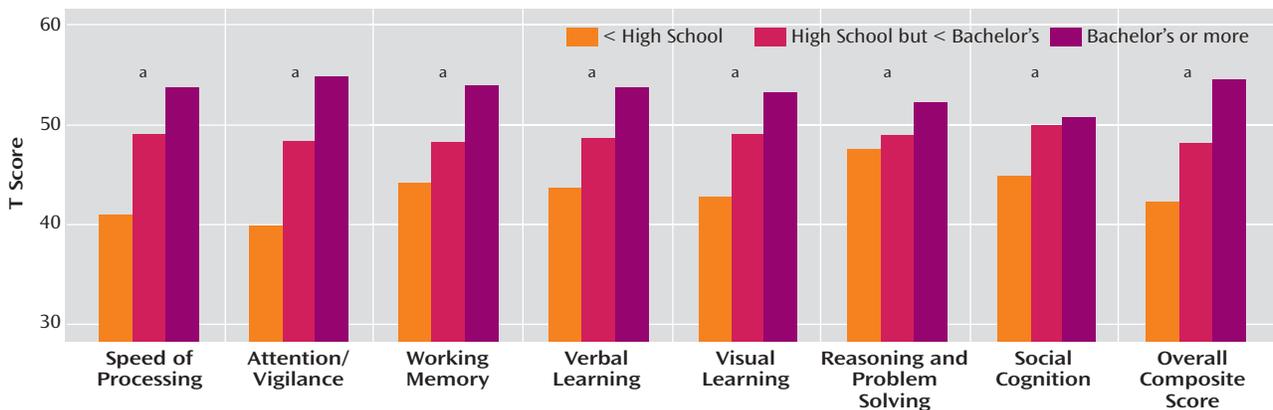
MCCB Scoring Program: Consideration of Age, Gender, and Education Effects

The data from the standardization study were used to develop a computer-based scoring program for the

MCCB. The program allows for three scoring options: no demographic correction; age and gender correction; and age, gender, and education correction. In applications of the MCCB to schizophrenia, effects of age and gender on cognition are typically sources of noise (nuisance variables) that may make the cognitive deficits of schizophrenia and changes in these deficits somewhat harder to detect. On the other hand, because schizophrenia often affects educational achievement, educational level in schizophrenia is not a simple nuisance variable, as it may partially reflect the severity of schizophrenia symptoms. Therefore, age and gender correction is the recommended normative scoring method for the MCCB in clinical trials of schizophrenia and is the default scoring option in the MCCB computer scoring program.

Development of the computer scoring program proceeded through a series of steps. Initially, the 300 cases from the community sample were divided into subgroups on the basis of age and gender (e.g., males 20–24 years old). Because our recruitment intentionally oversampled

FIGURE 3. Standardization Sample's Mean T Scores on the Cognitive Domains of the MATRICS Consensus Cognitive Battery, by Educational Level



^a Significant group effect, $p < 0.05$.

older individuals (two-thirds of the sample was age 40 and older), cases were then weighted (24) to achieve a sample representative of that reported for the 2000 U.S. census. Means and SDs were calculated for each test's primary score using the weighted sample.

The age and gender correction of scores from the MCCB scoring program employs a regression-based approach (25) using the case-weighted data from the standardization sample. Age and gender were used to predict each primary test score using a linear effects model. Although some more complex curvilinear age effects were noted for selected measures, correction for linear effects of age was judged to yield the most interpretable and generalizable procedure. Each participant's predicted score based on age and gender is therefore subtracted from the actual score to yield a residual score. The residual scores are then transformed to T scores (mean=50, SD=10). To maintain the composite scores on the same metric scale as the individual tests, T scores for the tests making up the composite are summed. The mean and SD of this sum are used to derive a T score with a mean of 50 and an SD of 10 for the composite. This method is used for deriving the T score values for the domains of working memory and speed of processing as well as for the overall composite score.

Discussion

The MATRICS Psychometric and Standardization Study marked the final step toward development of the MCCB. The primary aim was to develop a normative database using a single sample with the tests administered together as a unit. Given that the MCCB is a hybrid battery comprising independently developed tests, each with its own norms, this database serves as the common normative reference point for the derivation of cognitive domain and composite scores for use in future studies. Such a database was viewed within the MATRICS consensus process as particu-

larly important for a battery that will be a standard for assessing cognitive change in clinical trials.

The associations of MCCB test performance with age, gender, and education within the standardization study followed a pattern that made sense in light of prior research. Age and education effects were evident across most cognitive domains, with lower cognitive performance associated with increasing age and lower education (26). Gender effects were more varied, with differences reaching statistical significance in only three of the seven cognitive domains. Men performed better than women in the domains of reasoning and problem solving and of working memory, and women performed better in verbal learning. Given that the MCCB measure of reasoning and problem solving is a mazes test, the results are generally consistent with prior research indicating that, on average, males perform somewhat better on some visuospatial tasks and women on some verbal tasks (27).

With co-norming, a true integration of scores of the different tests and different cognitive domains in the battery can be achieved. Co-norming makes it possible to construct a profile of performance across cognitive domains, examine the relationships among scores of different tests, and further explore the cognitive abilities that are measured by this battery as a whole. These advantages for score interpretation should be understood in the context of the development of the battery. The seven cognitive domains included in the MCCB were settled on before test selection (28), and the MATRICS Neurocognition Committee decided which tests belonged to the selected cognitive domains. Consideration was given to administration time, and the number of tests included in the final battery was deliberately kept small. Although confirmatory factor analyses or other factor-analytic approaches may be used to test further whether, as prior studies suggest, seven cognitive factors are separable in schizophrenia, the MCCB alone would not be well suited to address this empirical question because several hypothesized cognitive factors

would have only one representative test. Perhaps a more productive approach would be to conduct confirmatory factor analyses with the MCCB plus other good representatives of the seven cognitive domains.

Some attention to the representativeness and limitations of our sample is warranted. First, the sample was of modest size compared with some normative samples, and it was deliberately unbalanced in the distribution of participants across the selected age groups. Age correction was best served by having more participants at ages for which cognitive changes were expected to be greatest, but this approach led to a less representative sample. To compensate for the unequal representation of persons across age groups, cases were weighted to approximate the age distribution reported for the 2000 U.S. census. Although case weighting addresses the distribution concern, it is not a substitute for actual cases. Second, the sample was not stratified by ethnicity to attain proportional representation of major ethnic and racial groups within each age or education bracket. Third, although the sites were selected in part to provide regional diversity, not all geographic regions of the United States were represented. The applicability of these normative results to populations in other countries is, of course, unknown. Also, because the data were collected at major universities and research centers, the normative sample was largely drawn from urban areas.

The MCCB is the result of the best efforts of a large group of individuals who participated in the MATRICS Initiative. Nonetheless, it represents judgments based on data and measures available at the time. The MCCB is based largely on standard neuropsychological instruments with well-established psychometric properties that favor their use in this battery and that have been reaffirmed in the present study. At the same time, advances in cognitive neuroscience over the past two decades have led to the development of new behavioral methods for assessing cognitive function and underlying neural mechanisms implicated in schizophrenia. A priority for future research will be to subject the new instruments emerging from basic cognitive neuroscientific research to rigorous psychometric testing and translate the most promising of them into clinically useful instruments for the evaluation of cognitive functioning in schizophrenia, possibly with improved sensitivity. Such efforts are under way in the NIMH-sponsored Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (<http://cntrics.ucdavis.edu/index.shtml>).

Received Jan. 7, 2007; revisions received July 7 and Sept. 5, 2007; accepted Sept. 11, 2007 (doi: 10.1176/appi.ajp.2007.07010043). From the Department of Psychology and the Semel Institute for Neuroscience and Human Behavior, Geffen School of Medicine, University of California, Los Angeles; VA Greater Los Angeles Healthcare System, Los Angeles; University of Kansas School of Medicine, Wichita; Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland Baltimore, Baltimore; NIMH, Bethesda, Md.; Department of Psychiatry and Behavioral Sciences, Duke University, Durham, N.C.; Massachusetts Mental Health Center, Division of Pub-

lic Psychiatry, Beth Israel Deaconess Medical Center, Boston; and the Department of Psychiatry, Harvard Medical School, Boston. Address correspondence and reprint requests to Dr. Kern, VA Greater Los Angeles Healthcare Center, MIRECC 210 A, Bldg. 210, Rm. 116, 11301 Wilshire Blvd., Los Angeles 90073; rkern@ucla.edu (e-mail).

Dr. Kern receives financial support from MATRICS Assessment, Inc., a nonprofit company formed after selection of the final MATRICS Consensus Cognitive Battery to allow its publication. Drs. Nuechterlein and Green have leadership positions but no financial interest in MATRICS Assessment, Inc. Drs. Keefe and Gold were developers of, and receive royalties from, the Brief Assessment of Cognition in Schizophrenia. Dr. Gold was a developer of, but has no financial interest in, the Letter-Number Span test. Dr. Nuechterlein has received research funding from Janssen LLP. Dr. Green has consulted for Abbott, Astellas, Bristol-Myers Squibb, Eli Lilly, Lundbeck, Memory Pharmaceuticals, Otsuka, Pfizer, Roche, Sanofi-Aventis, and Solvay. Dr. Keefe reports having received research funding, consulting fees, advisory board payments, or lecture or educational fees from Abbott, Acadia, AstraZeneca, Bristol-Myers Squibb, Cephalon, Dainippon Sumitomo Pharma, Eli Lilly, Forest Labs, Gabriel Pharmaceuticals, GlaxoSmithKline, Johnson & Johnson, Lundbeck, Memory Pharmaceuticals, Merck, NIMH, Orexigen, Otsuka, Pfizer, Repligen, Saegis, Sanofi-Aventis, and Xenoport. Dr. Marder has served as a consultant for Bristol-Myers Squibb, Otsuka, Pfizer, Merck, Roche, Solvay, and Wyeth. All other authors report no competing interests.

Dr. Fenton died in September 2006.

Funding for the MATRICS Initiative was provided through NIMH contract N01MH22006 to the University of California, Los Angeles (Dr. Marder, principal investigator; Dr. Green, co-principal investigator; Dr. Fenton, project officer). Funding for this study came from an option (Dr. Green, principal investigator; Dr. Nuechterlein, co-principal investigator) to the NIMH MATRICS Initiative.

The authors are grateful to the research staff at the five performance sites for their tireless effort on this project, Bi-Hong Deng for her coordinating assistance, Kelly Tillery for assistance in preparing the manuscript, and the UCLA Semel Institute Biostatistical Unit for developing the data entry systems.

References

- Green MF, Nuechterlein KH: The MATRICS initiative: developing a consensus cognitive battery for clinical trials. *Schizophrenia Res* 2004; 72:1–3
- 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-MATRICES conference to select cognitive domains and test criteria. *Biol Psychiatry* 2004; 56:301–307
- 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
- Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, Essock S, Fenton WS, Frese FJ III, Gold JM, Goldberg T, Heaton RK, Keefe RSE, Kraemer H, Mesholam-Gately R, Seidman LJ, Stover E, Weinberger DR, Young AS, Zalcman S, Marder SR: The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am J Psychiatry* (published online Jan 2, 2008; doi: 10.1176/appi.ajp.2007.07010042)
- Sullivan K, Bowden SC: Which tests do neuropsychologists use? *J Clin Psychol* 1997; 53:657–661
- Sweet JJ, Moberg PJ, Suchy Y: Ten-year follow-up survey of clinical neuropsychologists, part 1: practices and beliefs. *Clin Neuropsychol* 2000; 14:18–37
- Palmer BW, Appelbaum MI, Heaton RK: Rohling's interpretive method and inherent limitations on the flexibility of "flexible batteries." *Neuropsychol Rev* 2004; 14:171–176

8. Tulskey DS, Price LR: The joint WAIS-III and WMS-III factor structure: development and cross-validation of a six-factor model of cognitive functioning. *Psychol Assess* 2003; 15:149–162
9. Russell EW, Russell SL, Hill BD: The fundamental psychometric status of neuropsychological batteries. *Arch Clin Neuropsychol* 2005; 20:785–794
10. Kern RS, Green MF, Nuechterlein KH, Deng B: NIMH-MATRICES survey on assessment of neurocognition in schizophrenia. *Schizophr Res* 2004; 72:11–19
11. US Census Bureau: 2000 Census of Population and Housing: Profiles of General Demographic Characteristics, May 2001. <http://www.census.gov/prod/cen2000/dp1/2kh00.pdf>
12. Craik FI, Bialystok E: Cognition through the lifespan: mechanisms of change. *Trends Cogn Sci* 2006; 10:131–138
13. Faraone SV, Seidman LJ, Kremen WS, Pepple JR, Lyons MJ, Tsuang MT: Neuropsychological functioning among the non-psychotic relatives of schizophrenic patients: a diagnostic efficiency analysis. *J Abnorm Psychol* 1995; 104:286–304
14. Army Individual Test Battery: Manual of Directions and Scoring. Washington, DC, War Department, Adjutant General's Office, 1944
15. Keefe RSE: Brief Assessment of Cognition in Schizophrenia (BACS) Manual—A, Version 2.1. Durham, NC, Duke University Medical Center, 1999
16. Brandt J, Benedict RHB: The Hopkins Verbal Learning Test, Revised: Professional Manual. Odessa, Fla, Psychological Assessment Resources, Inc, 2001
17. Wechsler D: Wechsler Memory Scale, 3rd ed. San Antonio, Tex, Psychological Corporation, 1997
18. 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
19. White T, Stern RA: Neuropsychological Assessment Battery: Psychometric and Technical Manual. Lutz, Fla, Psychological Assessment Resources, Inc, 2003
20. Benedict RHB: Brief Visuospatial Memory Test, Revised: Professional Manual. Odessa, Fla, Psychological Assessment Resources, Inc, 1997
21. Spreen O, Strauss E: A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, 2nd ed. New York, Oxford University Press, 1998
22. Mayer JD, Salovey P, Caruso DR: Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT): User's Manual. Toronto, Multi-Health System Publishers, 2002
23. 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
24. Horton NJ, Kleinman KP: Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat* 2007; 61:79–90
25. Heaton RK, Avitable N, Grant I, Matthews CG: Further cross-validation of regression-based neuropsychological norms with an update for the Boston Naming Test. *J Clin Exp Neuropsychol* 1999; 21:572–582
26. Schaie KW: The course of adult intellectual development. *Am Psychol* 1994; 49:304–313
27. Halpern DF: Sex differences in intelligence. *Am Psychol* 1997; 52:1091–1102
28. Nuechterlein KH, Barch DM, Gold JM, Goldberg TE, Green MF, Heaton RF: Identification of separable cognitive factors in schizophrenia. *Schizophr Res* 2004; 72:29–39