

Evaluation of Functionally Meaningful Measures for Clinical Trials of Cognition Enhancement in Schizophrenia

Michael F. Green, Ph.D.

Nina R. Schooler, Ph.D.

Robert S. Kern, Ph.D.

Fred J. Frese, Ph.D.

Wendy Granberry, Pharm.D.

Philip D. Harvey, Ph.D.

Craig N. Karson, M.D.

Nancy Peters, R.N., M.B.A.

Michelle Stewart, Ph.D.

Larry J. Seidman, Ph.D.

John Sonnenberg, Ph.D.

William S. Stone, Ph.D.

David Walling, Ph.D.

Ellen Stover, Ph.D.

Stephen R. Marder, M.D.

Objective: Because reduction of psychotic symptoms in schizophrenia does not result in adequate community functioning, efforts have shifted to other areas, such as cognitive impairment. The U.S. Food and Drug Administration requires that drugs for cognition enhancement in schizophrenia show improvement on two distinct outcome measures in clinical trials: an accepted cognitive performance battery and a functionally meaningful coprimary measure. The authors examined the reliability, validity, and practicality of functionally meaningful measures.

Method: In this four-site validation study, schizophrenia patients were assessed at baseline (N=166) and 4 weeks later (N=144) on performance-based (Independent Living Scales, Test of Adaptive Behavior in Schizophrenia [TABS], and UCSD Performance-based Skills Assessment [UPSA]) and interview-based (Cognitive

Assessment Interview and Clinical Global Impression Scale for Cognition) candidate coprimary measures. In addition, cognitive performance, community functioning, and clinical symptoms were assessed. Both full and short forms of the performance-based measures were evaluated.

Results: All measures were well tolerated by patients, had adequate test-retest reliability, and showed good utility as a repeated measure. Measures differed in their correlation with cognitive performance, with performance-based measures having stronger correlations than interview-based measures. None of the measures had notable floor or ceiling effects or missing data.

Conclusions: Among the full-form measures, the UPSA was judged to have the strongest overall properties. Among the short forms, the TABS and UPSA appeared to have the strongest features. Use of the short forms saves time, but at the cost of lower test-retest reliability and weaker correlations with cognitive performance.

(*Am J Psychiatry* 2011; 168:400–407)

Enhancement of cognition in schizophrenia has become a major public health goal and drug development challenge. It has become increasingly clear that control of psychotic symptoms alone is not sufficient for community adaptation in schizophrenia, and efforts have been made to address unmet needs, including improvements in negative symptoms and cognitive impairment. To stimulate the development of cognition-enhancing drugs for schizophrenia, the National Institute of Mental Health (NIMH) created the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative (1–3). One key product of the MATRICS initiative was the selection of a standard cognitive battery for use in clinical trials: the MATRICS Consensus Cognitive Battery (MCCB) (4–6).

Typically the U.S. Food and Drug Administration (FDA) requires that a new drug improve a single outcome measure to receive approval for marketing. However, there are

specific situations in which a single outcome measure is considered insufficient because the clinical endpoint is multifaceted. In such cases, approval may depend on the drug improving two complementary measures that reflect different aspects of treatment response. These complementary measures are referred to as “coprimary outcomes.” The FDA has applied this coprimary requirement to drugs for improvement of cognition in schizophrenia: a drug for this use must show improvement both in cognitive performance and on a functionally meaningful measure in clinical trials (1). The FDA decided that measurement of cognitive performance with an accepted performance battery could not serve as a sole endpoint because it did not have an obvious and intuitive connection to improved overall outcome.

Although the endpoint for cognitive performance went through an extensive consensus process (2, 3), the FDA did not provide firm guidance on the definition of a

This article is featured in this month's AJP **Audio** and is the subject of a **CME** course (p. 445)

functionally meaningful coprimary outcome measure. Community functioning is a logical coprimary measure, but it is not practical because it is unlikely to change in the course of a typical clinical trial. In the context of clinical trials of cognition-enhancing drugs, practical coprimary measures might be intermediate between cognitive performance and daily functioning—for example, performance-based simulations of daily activities or interviews for cognition (see below). The FDA requirement for a coprimary measure presents a challenge to the field because of the absence of validated measures for this purpose.

MATRICES Coprimary and Translation (MATRICS-CT) is an NIMH initiative to further facilitate development of pharmacological agents for cognitive impairments in schizophrenia (7). A partnership of pharmaceutical companies supports it through donations to the Foundation at National Institutes of Health to address two remaining issues from the MATRICS initiative. One is to evaluate potential coprimary measures using a consensus process and collection of empirical data; these findings are summarized in this article. The second is to develop and validate high-quality translations of the MCCB for international trials (see <http://www.matricsinc.org>).

Potential coprimary measures use two approaches: performance-based and interview-based assessments. Performance-based (also called functional capacity) measures assess capacity to perform key tasks of daily living by asking participants to simulate real-world activities such as holding a social conversation, selecting grocery items to prepare a meal, and planning a trip using public transportation (8, 9). Good performance on such measures means that the person has the ability to perform the task, but not necessarily that they will perform the task in the community. Interview-based approaches ask people to estimate their cognitive abilities or the extent to which their daily lives are affected by cognitive impairment. Recently, cognitive assessment interviews have been developed specifically for patients with schizophrenia (10, 11).

An initial examination of coprimary measures was conducted in the MATRICS Psychometric and Standardization Study (PASS). The coprimary measures included in that study were suggested by the MATRICS Outcome Committee (A. Bellack, chair) but did not undergo a systematic selection process, as that was not the primary goal (12). MATRICS-CT sponsored a validation study called the Validation of Intermediate Measures (VIM) Study specifically to evaluate potential coprimary measures that were systematically selected. The study was designed and overseen by the MATRICS-CT VIM Committee, the members of which are coauthors of this article. The study aims were to examine 1) the psychometric characteristics of selected coprimary measures (test-retest reliability, interrater reliability, and utility as a repeated measure), 2) the validity of the measures (correlation with cognitive performance and community functioning), and 3) the measures'

practicality and tolerability (ease of setup, tester training and scoring, amount of missing data, assessment duration, and subject satisfaction ratings).

Method

Study Design and Participants

Clinical interviews were administered to determine eligibility and to assess cognitive performance and community functioning at baseline. Assessments on all coprimary measures and clinical symptoms were completed at baseline and at a 4-week follow-up.

To be eligible, participants had to be outpatients 18–60 years of age with a DSM-IV diagnosis of schizophrenia based on a diagnostic interview for this study or a previous one; have an understanding of spoken English adequate to comprehend testing procedures; have the ability to comprehend the consent form; and not have previously received the performance-based intermediate measures in this study, the MCCB, or similar cognitive assessment, within 6 months of study entry. Participants had to be clinically stable, as indicated by having no significant psychotropic medication changes in the past 2 months and none anticipated for the next month; showing evidence of stable symptomatology for at least 3 months; having Positive and Negative Syndrome Scale (PANSS) scores ≤ 4 (moderate) on P1 (delusions), P2 (conceptual disorganization), P3 (hallucinatory behavior), P5 (grandiosity), P6 (suspiciousness), and G8 (unusual thought content); having a PANSS score ≤ 15 on the negative symptoms subscale; and showing evidence that mood symptoms, if present, had been stable for at least 3 months.

Exclusion criteria were alcohol or other substance dependence in the past 6 months; alcohol or other substance abuse in the past 3 months; clinically significant neurological disease; head injury with loss of consciousness for more than 1 hour; a current medical condition that would interfere with valid assessment; dystonia or parkinsonism that would affect the validity of assessment; pregnancy or nursing; and current use of clozapine, potentially procognitive medications, antimentia medications, amphetamine, lithium, monoamine oxidase inhibitors, or tricyclic antidepressants. No benzodiazepines, sedatives, or anticholinergic medications were administered within 12 hours of assessment. After receiving a complete description of the study, participants provided written informed consent.

Sites

Four sites were selected by the VIM Committee. Each site had extensive experience in conducting schizophrenia clinical trials and local expertise in cognitive and performance-based assessment. Two academic sites (UCLA/Greater Los Angeles VA Healthcare System, the coordinating site in Los Angeles, and Harvard/Beth Israel Deaconess in Boston) and two freestanding clinical trial sites (Collaborative Neuroscience Network in Garden Grove, Calif., and Uptown Research Institute in Chicago) were selected.

Selection of Coprimary Measures

The selection process for study measures was modeled on the consensus and RAND panel process used by MATRICS for MCCB measure selection (see Figure S1 in the data supplement that accompanies the online edition of this article). Briefly, a MATRICS-CT subcommittee (K. Nuechterlein, chair) determined key criteria for selection of coprimary measures. Nominations were solicited broadly through announcements and e-mailings for measures that were categorized either as performance- or interview-based. The VIM Committee selected a subset of nominated measures for further evaluation. A comprehensive database of selected measures was developed by UCLA staff according to the evaluation criteria. The VIM Committee convened a RAND panel

meeting in February 2008 (13) to review the database and make recommendations for the VIM study. The RAND panelists (see acknowledgments) were selected as excellent representatives of their respective areas of expertise and for absence of any conflict of interest with the coprimary measures under consideration. Based on the RAND panel ratings and discussion, the VIM Committee selected the study measures described below.

Performance-Based Measures

Independent Living Scales (ILS). The ILS assesses adults' competence in instrumental activities of daily living (14). The items, which target situations relevant to independent living, require the examinee to solve problems, demonstrate knowledge, or perform a task. The ILS includes 70 items in five subscales: memory/orientation, managing money, managing home and transportation, health and safety, and social adjustment. The test yields two factors: problem solving, comprising primarily items that require knowledge of relevant facts, abstract reasoning, and problem solving ability; and performance/information, comprising primarily items that require general knowledge, short-term memory, and performance of simple, everyday tasks. The full scale score is a standardized score with a mean of 100 and a standard deviation of 15.

Test of Adaptive Behavior in Schizophrenia (TABS). The TABS includes five test areas (medication management, empty bathroom, shopping skills, clothes closet, and work and productivity) and one observed area (social skills) to assess skills needed for daily functioning (15). It focuses on initiation and problem identification. Props are used, such as pill containers in the medication management component and doll clothing in the clothes closet component. TABS scores are calculated as percent correct for each area; thus, the scores range from 0 to 100 per area. The total score is the mean of the six areas.

UCSD Performance-Based Skills Assessment (UPSA). The UPSA was designed to assess ability to perform everyday tasks needed for independent community functioning (16, 17). The UPSA evaluates five areas: household chores, communication, finance, transportation, and planning and recreational activities. It uses role play tasks that are administered as simulations of events that the person may encounter in the community. Raw scores from each subtest are transformed to yield comparable scores (ranging from 0 to 20) for each and a summary score ranging from 0 to 100 (higher scores reflect better performance).

Short Forms. The three performance-based measures were evaluated in their full forms. In addition, the TABS and UPSA had a short form, and the subtests that comprised the short forms were administered first to allow separate evaluation. The TABS short form included the medication and work and productivity subtests; the UPSA short form included the communication and finance components. Administering the short forms of these tests saves an estimated 15 minutes. The ILS does not have an identifiable short form, but the instrument yields two factor scores (performance and problem solving) that were evaluated separately.

Interview-Based Measures

Cognitive Assessment Interview (CAI). The CAI (18) is derived from two interview-based instruments, the Clinical Global Impression Scale for Cognition (11) and the Schizophrenia Cognition Rating Scale (10). Item response theory was used to select the items that provided the most information about the latent construct of interest, namely, interview-assessed cognitive deficit. From the original 41 items in both scales, 10 were selected that performed best across analyses, at several levels of the cognitive deficit construct, and showed good internal consistency. The CAI includes items that assess six of the seven MATRICS cognitive domains (all except visual learning). Items are rated on a 7-point

scale with defined anchor points referenced to healthy people of similar educational and sociocultural background. Higher scores reflect greater cognitive deficits that affect everyday functioning and/or greater need for support in performing those functions. In addition to the total score (the sum of the 10 items), which was the dependent measure in this study, the CAI includes a global rating of cognition (on a 100-point scale) similar to a Global Assessment of Functioning score (19).

Clinical Global Impression Scale for Cognition (CGI-Cognition). In an experimental component of the VIM study, a single-item 7-point scale was included to assess whether clinical raters can reliably rate cognitive impairment solely on the basis of a clinical symptom interview. The CGI-Cognition was modeled on the commonly used Clinical Global Impressions scale for symptom severity and did not include a manual or detailed anchor points.

Additional Measures

The VIM study included three additional measures. The MCCB global composite score (4, 5) was included to examine the relationship of the coprimary measures to cognitive performance. The Heinrichs-Carpenter Quality of Life Scale summary score (20) was used to examine relationships with community functioning. This score excluded seven intrapsychic items that assess negative symptoms. The PANSS was used to evaluate psychopathology (21).

Training on each of the coprimary measures was conducted at an in-person start-up meeting and in subsequent teleconferences. Trainees had prior experience working with schizophrenia patients. Training on the CAI, TABS, and UPSA was conducted by a developer of the measure. ILS training was provided by an investigator with extensive experience with that measure.

To ensure independence of the ratings, each site used at least three raters. One determined study eligibility, collected demographic information, and administered the PANSS, CGI-Cognition, and Quality of Life Scale at baseline and the PANSS and CGI-Cognition at week 4. A second completed the performance-based intermediate measures at both assessment points and the MCCB at baseline, and a third rater completed the CAI at both assessment points. Assessments were completed with information from participants only; although informants' ratings can be included in the CAI and Quality of Life Scale, they were not in this study to approximate the likely conditions of multisite clinical trials. The performance-based measures were administered in a counterbalanced order to allow examination of order effects.

Results

A total of 196 participants gave consent and were screened. Of these, 166 received baseline assessments (27 were ineligible; three withdrew consent). Three patients were excluded because of invalid data, leaving 163 participants with valid baseline assessments. Of these, 144 (88.3%) were tested at the 4-week follow-up.

The sample was comparable to other clinically stable samples in schizophrenia trials in gender, age, education, and symptom severity (22, 23). The sample was about one-third female (N=58, 35.6%), with a mean age of 43.9 years (SD=10.1) and a mean of 12.3 years of education (SD=2.1). The mean duration of illness was 20.3 years (SD=10.6, range=1–43). Participants were clinically stable over the course of the study; the mean total PANSS score was 61.5 (SD=12.6) at baseline and 61.8 (SD=13.2) at 4 weeks.

TABLE 1. Descriptive Data From Candidate Coprimary Measures for Clinical Trials of Cognition Enhancement in Schizophrenia Administered at Baseline to a Sample of Patients With Schizophrenia

Measure	N	Mean	SD	Range	Possible Range
Independent Living Scales	163	88.0	15.1	55–118	55–121
Test of Adaptive Behavior in Schizophrenia	160	69.3	10.1	46.8–91.3	0–100
UCSD Performance-Based Skills Assessment	162	73.0	11.6	46.1–98.3	0–100
Clinical Global Impression Scale for Cognition ^a	163	3.05	0.89	1–5	1–7
Cognitive Assessment Interview ^a	163	29.83	9.38	10–53	10–70
MATRICES Consensus Cognitive Battery	160	27.9	11.4	1–62	1–100
Quality of Life Scale	160	37.3	13.4	3–70	0–78

^a Higher scores indicate greater impairment.

The summary data for the main variables are presented in Table 1. All scores were inspected for range and distribution, and no transformations were deemed necessary. The MCCB score indicates that the sample was slightly more than two standard deviations below the mean for age- and gender-corrected norms, comparable to the MATRICES PASS study (24). There were no significant order effects.

Key Scientific Criteria

In evaluating the measures, an a priori decision was made to prioritize test-retest reliability and correlation with cognitive performance, characteristics that are considered to be most important for a coprimary measure in trials of cognition-enhancing drugs. Test-retest reliability translates directly into power estimates and sample size requirement, and it was considered the most important test characteristic during the MATRICES initiative (3). Correlation with cognitive performance was also considered essential given the definition of a coprimary measure: for a drug to receive FDA approval for cognition enhancement, it has to significantly improve both cognitive performance and the coprimary measure. Hence, correlation of the two outcome measures is viewed as desirable so that a drug would not need to affect two independent constructs.

Psychometrics and validity data are presented in Table 2. Regarding test-retest reliability, 0.70 is a conventional cutoff for acceptability for measures of this type. The ILS, CAI, and UPSA had intraclass correlation coefficients (ICCs) above this value, and the TABS and CGI-Cognition just below (ICC=0.69). For the short forms, test-retest reliability values were all very close to 0.70. In addition to test-retest reliability, interrater agreement is an important consideration for the CAI. Interrater reliability (ICC=0.73) was determined on the basis of a set of eight tapes that were rated by all study CAI raters.

For correlation with cognitive performance, there were considerable differences among the full measures. Overall, the performance measures had much greater overlap with cognitive performance (up to 45% shared variance with the UPSA) than the interview-based measures (5% and 14% shared variance for the CAI and CGI-Cognition, respectively). Paired contrasts are reported in a footnote to Table 2. For the short forms, three of the measures were identical (28% shared variance); the ILS problem solving component was lower (15%).

Additional Scientific Criteria

Table 2 also presents data for utility as a repeated measure and relationship to community functioning. All of the measures had small to modest practice effects (the largest effect sizes were 0.24 for both full measures and short forms). The number of scores at floor or ceiling was not considered to be problematic for any measure at either testing time. Correlation with community functioning was generally low for all measures. The CGI-Cognition and the short forms of the TABS and UPSA had the lowest correlations, in the range of 0.12–0.15. Other measures were in the 0.23–0.30 range.

Practicality and Tolerability

Data on practicality and tolerability were collected only on full measures (Table 3), and they were not collected on the CGI-Cognition, which is a single-item measure. Practicality and tolerability were each rated on a 7-point scale, with higher scores indicating better ratings. Practicality was rated by five testers for ease of set-up, administration, and scoring. Study participants rated tolerability based on how pleasant or unpleasant they found the test to be. Practicality showed a range across measures, with the UPSA scoring the best and the CAI the poorest. All of the measures were well tolerated by patients, with mean scores ranging from 5.4 to 6.0. The administration times differed notably across measures; the ILS took the longest time (46 minutes), and the CAI the shortest (25 minutes). Each administration time was statistically different from every other. The amount of missing data was very small for all measures (1.6% for the TABS, 0.6% for the UPSA, and none for the others, across both assessments).

Relationships to Symptoms

Table 4 presents the correlations with the five factors of the PANSS (25). In general, the correlations with positive and negative symptoms were relatively low. The largest correlations were seen with the disorganized thought factor. The measure with the largest correlation with clinical symptoms was the CGI-Cognition, which was completed by the PANSS rater.

Site Differences

One-way analyses of variance were conducted to examine site effects. There were no site differences in age and illness chronicity, but there were in education ($F=8.72$,

TABLE 2. Psychometric and Validity Data for Candidate Coprimary Measures for Clinical Trials of Cognition Enhancement in Schizophrenia Administered to a Sample of Patients With Schizophrenia (N=163)

Measure	Key Scientific Criteria					Additional Scientific Criteria			
	Test-Retest Reliability			Correlation With MCCB ^b		Utility as a Repeated Measure			Correlation With QLS ^c
	ICC ^a	r	95% CI	r	r ²	Floor/Ceiling at T1 (N)	Floor/Ceiling at T2 (N)	T1-T2 Effect Size	r
Full measures									
Independent Living Scales	0.76	0.77	0.69–0.83	0.51	0.26	4/0	2/0	0.15	0.30
Test of Adaptive Behavior in Schizophrenia	0.69	0.71	0.62–0.78	0.61	0.37	0/0	0/0	0.24	0.23
UCSD Performance-Based Skills Assessment	0.74	0.75	0.67–0.81	0.67	0.45	0/0	0/0	0.18	0.25
Clinical Global Impression Scale for Cognition	0.69	0.69	0.59–0.77	0.38	0.14	6/0	5/0	0.03	0.12
Cognitive Assessment Interview	0.76	0.77	0.69–0.83	0.23	0.05	1/0	1/0	0.10	0.27
Short forms									
Independent Living Scales, performance subtest	0.70	0.71	0.62–0.78	0.53	0.28	0/0	0/0	0.10	0.24
Independent Living Scales, problem solving subtest	0.68	0.68	0.58–0.76	0.39	0.15	9/0	6/0	0.07	0.27
Test of Adaptive Behavior in Schizophrenia	0.67	0.69	0.59–0.77	0.53	0.28	0/0	0/1	0.24	0.15
UCSD Performance-Based Skills Assessment	0.69	0.69	0.59–0.77	0.53	0.28	0/0	0/5	0.12	0.15

^a ICC=Intraclass correlation coefficient.

^b MCCB=MATRICES Consensus Cognitive Battery. Contrasts: UCSD Performance-Based Skills Assessment > Independent Living Scales, Cognitive Assessment Interview, Clinical Global Impression Scale for Cognition; Test of Adaptive Behavior in Schizophrenia > Cognitive Assessment Interview, Clinical Global Impression Scale for Cognition; Independent Living Scales > Cognitive Assessment Interview.

^c QLS=Quality of Life Scale.

TABLE 3. Ratings of Practicality and Tolerability of Candidate Coprimary Measures for Clinical Trials of Cognition Enhancement in Schizophrenia Administered to a Sample of Patients With Schizophrenia (N=163)^a

Measure	Practicality	Tolerability ^b	Duration ^c (Minutes)
Independent Living Scales	4.6	5.4	46
Test of Adaptive Behavior in Schizophrenia	5.2	5.9	33
UCSD Performance-Based Skills Assessment	6.0	5.6	27
Cognitive Assessment Interview	4.1	6.0	25

^a Practicality and tolerability were rated on 1–7 scale where 7=best.

^b Cognitive Assessment Interview, Test of Adaptive Behavior in Schizophrenia > UCSD Performance-Based Skills Assessment > Independent Living Scales.

^c Independent Living Scales > Test of Adaptive Behavior in Schizophrenia > UCSD Performance-Based Skills Assessment > Cognitive Assessment Interview.

df=3, 162, p<0.001) and PANSS total score (F=59.02, df=3, 162, p<0.001). Participants at the two academic sites were more educated and less symptomatic than those at the two freestanding sites. There were site differences on all coprimary measures (ILS: F=7.81, df=3, 163, p<0.001; TABS: F=4.38, df=3, 159, p=0.005; UPSA: F=5.86, df=3, 161, p=0.001; CAI: F=23.86, df=3, 162, p<0.001; CGI-Cognition: F=4.79, df=3, 162, p=0.003). Differences for the three performance-based measures were mainly due to better performance among the patients at UCLA; differences in the interview-based measures were primarily due to greater impairment among participants at Collaborative Neuroscience Network.

Discussion

In this four-site study, we examined the psychometrics, validity, and practicality of candidate performance- and interview-based measures for coprimary outcome in

clinical trials of cognition enhancement in schizophrenia. Assessments were conducted at baseline on 163 participants and on more than 88% at 4-week follow-up. Full measures and short forms were evaluated separately. The full forms of the performance-based measures performed the best overall.

Regarding the key scientific criteria, all of the full measures and short forms had acceptable test-retest reliability, with correlations around 0.70, and the ILS, CAI, and UPSA exceeded that threshold. For the interview-based CAI, interrater reliability was an additional source of variability. VIM study raters received extensive, ongoing training by one of the CAI developers, and the same rater completed both ratings 88% of the time. Studies that use the same rater less frequently should expect lower levels of reliability than those reported here.

Regarding the relationship of these measures to cognitive performance, there were significant and substantial differences among them, with the performance-based

TABLE 4. Correlations Between Candidate Coprimary Measures for Clinical Trials of Cognition Enhancement in Schizophrenia and the Five Factors of the Positive and Negative Syndrome Scale in a Sample of Patients With Schizophrenia (N=163)

Measure	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	Negative Symptoms	Positive Symptoms	Disorganized Thought	Uncontrolled Hostility/Excitement	Anxiety/Depression
Independent Living Scales	-0.17*	-0.21**	-0.23**	-0.08	-0.05
Test of Adaptive Behavior in Schizophrenia	-0.13	-0.13	-0.32**	-0.02	0.02
UCSD Performance-Based Skills Assessment	-0.17*	-0.23**	-0.33**	-0.06	-0.08
Clinical Global Impression Scale for Cognition ^a	0.21**	0.34**	0.43**	0.04	0.08
Cognitive Assessment Interview ^a	0.11	0.11	-0.05	-0.18*	0.02

^a Higher scores indicate greater impairment.

* $p < 0.05$. ** $p < 0.01$.

measures showing much more overlap (shared variance of 26%–45%) compared with the interview-based measures (5%–14%). With only 5% shared variance between the CAI and the MCCB, it appears that the CAI primarily measures a different construct than cognitive performance.

Differences among measures were not as pronounced for the secondary scientific criteria (utility as a repeated measure and community functioning), and all of the measures were well tolerated. Correlations with community functioning were relatively low for all measures, perhaps in part because of the absence of informant ratings for the Quality of Life Scale. The results also suggest that being able to perform an activity does not necessarily mean that the person does so in the community. The full measures differed substantially in administration time, with the ILS taking the longest.

Based on these data, the VIM Committee considered the UPSA to be the leading coprimary measure among the full measures because it had several strong features: good test-retest reliability, excellent shared variance with cognitive performance, good utility as a repeated measure with no problematic floor or ceiling effects, and reasonable tolerability and practicality. Among the short forms, three of the measures performed comparably across the criteria (the TABS, the UPSA, and the ILS performance factor). The committee considered the TABS and UPSA short forms to have an advantage over the ILS performance factor because the short forms are self-contained. The ILS performance factor includes items administered throughout the test, and it is not known whether the same psychometric properties and validity would be obtained if those items were administered without the rest of the test.

The study findings reveal the inherent trade-offs in using short forms of the coprimary measures. Although these short forms save some administration time, they have lower reliability and lower shared variance with cognitive functioning. For the UPSA, the difference in reliability between short form and full measure (ICC=0.69 compared with ICC=0.74) could translate into a notable difference in the sample size required for adequate statistical power, depending on the study design and covariates. Similarly, the difference in shared variance with cognitive performance (28% compared with 45%) influences the confidence that

improvement in cognitive performance would be accompanied by improvement in the coprimary measure.

One limitation of this study is that it was conducted in the United States and used English-language versions of the tests. The RAND panel commented that some nominated coprimary tests would be particularly hard to adapt in other cultures (e.g., those that involve videotaped stimuli or specific skills training components), and these tests were not included in the VIM study. There are data indicating that modified versions of coprimary tests perform similarly across different Western cultures (26), but we do not know the extent to which these measures will need modification for broader international applications. A separate component of MATRICS-CT conducted international surveys with clinicians to start to address the question of global adaptation. Another limitation is that the study was not a treatment trial and therefore was not designed to assess the sensitivity of coprimary measures to change in the context of cognitive enhancement. Nonetheless, the VIM study provides psychometric and validity data on the coprimary measures that suggest the likelihood of detecting underlying changes in cognition when such changes occur.

Placed in the broader context of the MATRICS and MATRICS-CT initiatives, this study is an important methodological step in a pathway to FDA drug approval for critical unmet needs in schizophrenia. The rationale for these initiatives is that substantial gains in recovery from schizophrenia will require treatments for cognitive impairment and negative symptoms. The pathway for drug approval for cognition has specific requirements: a scientific consensus about definition and measurement of cognitive performance and a coprimary measure with more face validity for patient improvement. Unlike the situation with cognitive performance measures, the FDA has not taken the position that a single coprimary measure be identified. The data from this study provide guidance for selection among currently available measures, as well as guidance for development of new coprimary measures in this rapidly developing area. From the clinician's perspective, the study serves as a reminder that schizophrenia patients who live in the community with relatively few and stable symptoms have substantial cognitive impairments

that are correlated with difficulties in performing clinically meaningful daily tasks, such as planning an outing to the park or using the telephone to make an appointment. Development of treatments for these impairments is an urgent need.

Presented at a meeting sponsored by NIMH for MATRICS-CT (Co-Primary and Translation), Bethesda, Md., October 27, 2009, and at the annual meeting of the American College of Neuropsychopharmacology, Hollywood, Fla., December 7, 2009. Received March 22, 2010; revisions received July 13, Sept. 1, Sept. 22, and Oct. 15, 2010; accepted Nov. 8, 2010 (doi: 10.1176/appi.ajp.2010.10030414). From the Semel Institute of Neuroscience and Human Behavior, UCLA, Los Angeles; VA Desert Pacific Mental Illness Research, Education, and Clinical Center, Los Angeles; SUNY Downstate Medical Center, Brooklyn, N.Y.; VISN 5 Mental Illness Research, Education, and Clinical Center, Baltimore; Northeastern Ohio Universities College of Medicine, Rootstown; GlaxoSmithKline, Research Triangle Park, N.C.; Psychiatry and Behavioral Sciences, Emory University, Atlanta; Merck, North Wales, Pa.; Sanofi-Aventis, Malvern, Pa.; Pfizer, Inc., New London, Conn.; Beth Israel Deaconess Medical Center/Harvard Medical School, Boston; Uptown Research Institute, Chicago; Collaborative Neuroscience Network, Garden Grove, Calif.; and NIMH, Bethesda, Md. Address correspondence and reprint requests to Dr. Green, 300 UCLA Medical Plaza, Rm. 2263, Los Angeles, CA 90095-6968; mgreen@ucla.edu (e-mail).

Dr. Green has served as a consultant for Abbott Laboratories, Astellas, Cypress, Dainippon Sumitomo Pharma, GlaxoSmithKline, Lundbeck, Otsuka, Sanofi-Aventis, Takeda, Teva, and Wyeth and has been a speaker for Janssen-Cilag, Otsuka, and Sunovion. Dr. Schooler has received research support or served as a consultant or speaker for Abbott Laboratories, AstraZeneca, Bristol-Myers Squibb, Dainippon Sumitomo Pharma, Eli Lilly, Hoffmann-La Roche, Janssen Cilag-Spain, Lundbeck, Merck, OrthoMcNeil Janssen, Pfizer, and Schering Plough. Dr. Kern has served as a consultant for Otsuka. Dr. Granberry is an employee of GlaxoSmithKline. Dr. Harvey has served as a consultant for Abbott, Cypress, Eli Lilly, Merck, Johnson & Johnson, Solvay, Dainippon Sumitomo Pharma America, and Wyeth and has received research support from AstraZeneca. Dr. Karson is an employee of Merck. Ms. Peters is an employee of Sanofi-Aventis. Dr. Stewart is an employee of Pfizer. Dr. Sonnenberg has provided clinical trial or consulting services for Abbott, Bristol-Myers Squibb, Cephalon, Dainippon Sumitomo Pharma, Eli Lilly, Enviro, Forest, Johnson & Johnson, Lundbeck, Novartis, Otsuka, Pfizer, Shire, Solvay, Sanofi-Aventis, and Takeda. Dr. Stone has received research support from Ortho-McNeil Janssen Scientific Affairs. Dr. Walling has provided clinical trial or consulting services to Abbott, Alexza, AstraZeneca, BrainCell, Cephalon, Dainippon Sumitomo Pharma, Eisai, Eli Lilly, Forest, GlaxoSmithKline, Johnson & Johnson, Lundbeck, Memory Pharmaceuticals, Novartis, Otsuka, Pfizer, Roche, Sepracor, Shire, Solvay, Sanofi-Aventis, Takeda, Toyama, Vanda, and Wyeth and has served as a consultant or speaker for Abbott, GlaxoSmithKline, Janssen, Novartis, and Otsuka. Dr. Marder has been a consultant for Wyeth, Otsuka, Pfizer, Schering Plough, Lundbeck, and Sanofi-Aventis. The other authors report no financial relationships with commercial interests.

Supported by an extension to contract HHSN 278 2004 41003C from NIMH, with additional support from grant MH042392.

The authors thank the members of the MATRICS-CT (Co-Primary and Translation) Scientific Board, which consisted of representatives from academia, the pharmaceutical industry, NIMH, and the Foundation at NIH. The board provided excellent input and guidance for this study. The authors also thank the members of the MATRICS-CT Rand Panel who are not listed as authors on this paper: Drs. Deanna Barch, John Brekke, Judith Cook, Patrick Corrigan, Michael Egan, Helena Kraemer, William Lawson, Andy Leon, Steve Romano, and Sophia Vinogradov.

References

- Buchanan RW, Davis M, Goff D, Green MF, Keefe RS, Leon AC, Nuechterlein KH, Laughren T, Levin R, Stover E, Fenton W, Marder SR: A summary of the FDA-NIMH-MATRICES workshop on clinical trial design for neurocognitive drugs for schizophrenia. *Schizophr Bull* 2005; 31:5–19
- 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
- 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
- Nuechterlein KH, Green MF: MATRICS Consensus Cognitive Battery. Los Angeles, Calif, MATRICS Assessment, Inc, 2006
- Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, Essock S, Fenton WS, Frese FJ 3rd, Gold JM, Goldberg T, Heaton RK, Keefe RS, 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* 2008; 165:203–213
- Kern RS, Nuechterlein KH, Green MF, Baade LE, Fenton WS, Gold JM, Keefe RS, 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* 2008; 165:214–220
- Stover EL, Brady L, Marder SR: New paradigms for treatment development. *Schizophr Bull* 2007; 33:1093–1099
- 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
- Bellack AS, Sayers M, Mueser K, Bennett M: Evaluation of social problem solving in schizophrenia. *J Abnorm Psychol* 1994; 103:371–378
- Keefe RSE, Poe M, Walker TM, Kang JW, Harvey PD: The Schizophrenia Cognition Rating Scale: an interview-based assessment and its relationship to cognition, real-world functioning, and functional capacity. *Am J Psychiatry* 2006; 163:426–432
- Ventura J, Cienfuegos A, Boxer O, Bilder R: Clinical Global Impression of Cognition in Schizophrenia (CGI-CogS): reliability and validity of a co-primary measure of cognition. *Schizophr Res* 2008; 106:59–69
- Green MF, Nuechterlein KH, Kern RS, Baade LE, Fenton WS, Gold JM, Keefe RS, 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* 2008; 165:221–228
- Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDonald J, Vader JP, Kahan JP: The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, Calif, RAND Institute, 2001
- Loeb PA: Independent Living Scales Manual. San Antonio, Tex, Psychological Corp, 1996
- Velligan DI, Diamond P, Glahn DC, Ritch J, Maples N, Castillo D, Miller AL: The reliability and validity of the Test of Adaptive Behavior in Schizophrenia (TABS). *Psychiatry Res* 2007; 151:55–66
- Patterson TL, Goldman S, McKibbin CL, Hughs T, Jeste DV: UCSD performance-based skills assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophr Bull* 2001; 27:235–245
- Twamley EW, Doshi RR, Nayak GV, Palmer BW, Golshan S, Heaton RK, Patterson TL, Jeste DV: Generalized cognitive impairments, ability to perform everyday tasks, and level of independence in community living situations of older patients with psychosis. *Am J Psychiatry* 2002; 159:2013–2020

18. Reise SP, Ventura J, Keefe RSE, Baade LE, Gold JM, Green MF, Kern RS, Mesholam-Gately R, Nuechterlein KH, Seidman LJ, Bilder RM: Bifactor and item response theory analyses of interviewer report scales of cognitive functioning in schizophrenia. *Psychol Assess* (in press)
19. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders, 4th ed (DSM-IV)*. Washington, DC, American Psychiatric Association, 1994
20. Heinrichs DW, Hanlon TE, Carpenter WT: The Quality of Life Scale: an instrument for rating the schizophrenic deficit syndrome. *Schizophr Bull* 1984; 10:388–398
21. Kay SR, Opler LA, Lindenmayer JP: The Positive and Negative Syndrome Scale (PANSS): rationale and standardization. *Br J Psychiatry* 1989; 155(suppl 7):59–65
22. Goff DC, Lamberti JS, Leon AC, Green MF, Miller AL, Patel J, Manschreck T, Freudenreich O, Johnson SA: A placebo-controlled add-on trial of the ampakine, CX516, for cognitive deficits in schizophrenia. *Neuropsychopharmacology* 2008; 33:465–472
23. Keefe RS, Malhotra AK, Meltzer HY, Kane JM, Buchanan RW, Murthy A, Sovel M, Li C, Goldman R: Efficacy and safety of donepezil in patients with schizophrenia or schizoaffective disorder: significant placebo/practice effects in a 12-week, randomized, double-blind, placebo-controlled trial. *Neuropsychopharmacology* 2008; 33:1217–1228
24. Kern RS, Gold JM, Dickinson D, Green MF, Nuechterlein KH, Baade LE, Keefe RSE, Mesholam-Gately R, Seidman LJ, Lee C, Sugar CA, Marder SR: The MCCB impairment profile for schizophrenia outpatients: results from the MATRICS Psychometric and Standardization Study. *Schizophr Res* (in press)
25. Marder SR, Davis JM, Chouinard G: The effects of risperidone on the five dimensions of schizophrenia derived by factor analyses: combined results of the North American trials. *J Clin Psychiatry* 1997; 58:538–546
26. Harvey PD, Helldin L, Bowie CR, Heaton RK, Olsson AK, Hjarthag F, Norlander T, Patterson TL: Performance-based measurement of functional disability in schizophrenia: a cross-national study in the United States and Sweden. *Am J Psychiatry* 2009; 166:821–827