

Functional Co-Primary Measures for Clinical Trials in Schizophrenia: Results From the MATRICS Psychometric and Standardization Study

Michael F. Green, Ph.D.

Keith H. Nuechterlein, Ph.D.

Robert S. Kern, 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.

Larry J. Seidman, Ph.D.

Ellen Stover, Ph.D.

Stephen R. Marder, M.D.

Objective: During the consensus meetings of the National Institute of Mental Health Measurement and Treatment Research to Improve Cognition in Schizophrenia (NIMH-MATRICES) Initiative, the U.S. Food and Drug Administration took the position that a drug for this purpose should show changes on 1) an accepted consensus cognitive performance measure and 2) an additional measure (i.e., a co-primary) that is considered functionally meaningful. The goal of the current study was to describe steps to evaluate four potential co-primary measures for psychometric properties and validity.

Method: As part of the five-site MATRICS Psychometric and Standardization Study (PASS), two measures of functional capacity and two interview-based measures of cognition were evaluated in 176 patients with schizophrenia (167 of these patients were retested 4 weeks later).

Results: Data are presented for each co-primary measure for test-retest reliability, utility as a repeated measure, relationship to cognitive performance, relationship to functioning, tolerability/practicality, and number of missing data.

Conclusions: Psychometric properties of all of the measures were considered acceptable, and the measures were generally comparable across the various criteria, except that the functional capacity measures had stronger relationships to cognitive performance and fewer missing data. The development and evaluation of potential co-primary measures is still at an early stage, and it was decided not to endorse a single measure for clinical trials at this point. The current findings offer the initial steps to identify functionally meaningful co-primary measures in this area and will help to guide further evaluation of such measures.

(*Am J Psychiatry* 2008; 165:221–228)

The National Institute of Mental Health Measurement and Treatment Research to Improve Cognition in Schizophrenia (NIMH-MATRICES) Initiative was created to stimulate the development of cognition-enhancing drugs for schizophrenia (1–3). As described in two other articles (4, 5), a key deliverable for the MATRICS Initiative was the selection, through a broad-based multidisciplinary consensus process, of a standard cognitive battery: the MATRICS Consensus Cognitive Battery (6). The U.S. Food and Drug Administration (FDA) indicated at the MATRICS meetings that significant improvement on a consensus cognitive performance endpoint would be necessary, but not sufficient, for drug approval. In addition to changes in cognitive performance, the FDA will require improvements on a functionally meaningful co-primary measure that would have more face validity for consumers and clinicians than cognitive performance measures (3). This requirement presents a notable challenge because of the absence of accepted or validated co-primary measures for this purpose.

One possible co-primary measure might be an assessment of community functioning. However, change in community status (e.g., return to work, increased social relationships, or higher degree of independent living) involves many intervening variables (both personal and social factors) that act between underlying cognitive processes and these functional outcomes. These intermediate steps would make it difficult to see the functional benefits of cognition-enhancing effects (7–10). Similarly, improvements in cognition would be expected to take considerable time to translate into functional improvements (11). Finally, changes in daily functioning would depend on nonbiological factors that are typically uncontrolled in clinical trial studies (e.g., the availability of psychosocial rehabilitation, social support networks, local employment rates, and training opportunities). Hence, alternative co-primary measures were considered that might change more directly and on a comparable time course with cognitive improvement. Such measures include standardized

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

TABLE 1. Test-Retest Reliability of Measures in the MATRICS Psychometric and Standardization Study

Co-Primary Measure	Test-Retest Reliability (intraclass correlation)
Maryland Assessment of Social Competence	
Conversational content	0.68
Nonverbal content	0.66
Effectiveness	0.69
UCSD Performance-Based Skills Assessment	
Total score	0.70
Medication management	0.48
Schizophrenia Cognition Rating Scale	
Overall impression (patient)	0.60
Overall impression (informant)	0.77
Overall impression (interviewer)	0.82
Clinical Global Impression of Cognition in Schizophrenia	
Neurocognitive state (patient)	0.76
Neurocognitive state (caregiver)	0.85
Neurocognitive state (composite)	0.80
Global Assessment of Functioning (composite)	0.89

tests of functional capacity or interview-based assessments of cognition.

Functional capacity refers to an individual's capacity for performing key tasks of daily living. To assess functional capacity, participants simulate in the clinic such real-world activities as holding a social conversation, preparing a meal, or taking public transportation (12, 13). Good performance on such measures does not mean that a person will perform the tasks in the community, but it does mean that the person could perform the task if he or she had the opportunity and was willing. Because performance on measures of functional capacity do not depend on social and community opportunities, they are more likely to be temporally linked with treatment-related changes in underlying cognition.

Another approach for co-primary measures was to consider interview-based assessments of cognitive abilities. Different interview-based approaches to cognition have been used, including asking people to estimate their own cognitive abilities or asking subjects to estimate the extent to which their daily lives are affected by cognitive impairment. Interview-based approaches present a challenge because it is often difficult for individuals (healthy subjects as well as patients) to estimate their own performance abilities (14, 15). Recently, some cognitive assessment interviews have been developed in which the ratings of psychotic patients are supplemented with ratings from informants (e.g., caregivers), an approach that might have advantages over previous assessments that used only self-reports (16, 17).

The goal of the current study was to evaluate the reliability, validity, and appropriateness for use in clinical trials of four potential co-primary measures: two measures of functional capacity and two interview-based measures of cognition. With this goal in mind, these four measures were added to the MATRICS Psychometric and Standardization Study (PASS) and assessed for their test-retest reliability, utility as a repeated measure, relationship to cognitive performance, relationship to outcome, practicality/ tolerability, and number of missing data.

ability, utility as a repeated measure, relationship to cognitive performance, relationship to outcome, practicality/ tolerability, and number of missing data.

Method

The complete methods for this study are presented online (the data supplement is available at <http://ajp.psychiatryonline.org>). A brief summary is presented here.

The data in this article are part of the five-site MATRICS PASS. Phase 1 was conducted with 176 schizophrenia patients tested twice at a 4-week interval (4), and phase 2 included 300 community subjects to collect co-norming data for the tests in the final battery (5). To evaluate the co-primary measures, we used a similar approach to that used to evaluate the cognitive performance measures and considered 1) test-retest reliability, 2) utility as a repeated measure, 3) relationship to functional status, 4) tolerability/practicality, and 5) number of missing data. In addition, we evaluated the degree to which co-primary measures correlated with cognitive performance (3).

Based on discussions at consensus meetings and recommendations of the MATRICS Outcomes Committee (1) (A. Bellack, chair), two approaches to co-primary measures were considered: measures of functional capacity and self-report measures of cognition. Two measures were selected from each approach. Unlike the large consensus and data collection process that was used to select cognitive performance tests for the MATRICS Consensus Cognitive Battery, the selection of potential co-primary measures was based on expert recommendations from the committee. The functional capacity measures were the Maryland Assessment of Social Competence (13) and the University of California at San Diego (UCSD) Performance-Based Skills Assessment (18). The interview-based measures of cognition were the Schizophrenia Cognition Rating Scale (16) and the Clinical Global Impression of Cognition in Schizophrenia (17).

Cognitive performance was assessed with the MATRICS Consensus Cognitive Battery (4, 5). As described more fully in the first article (4), community functioning was assessed with variables from the Birchwood Social Functioning Scale (19), supplemented with work and school items from the Social Adjustment Scale (20).

Results

Study Group

The study group for these analyses is the same as described in the first article (4): across the five performance sites, 176 patients were assessed at baseline and 167 were assessed at the 4-week follow-up.

Test-Retest Reliability

The results for test-retest reliability are shown in Table 1. The table shows the intraclass correlation coefficient (ICC), which takes into account changes in mean level. Both the ICC and Pearson's *r* are shown in the online data supplement; differences for the two statistics were minor. The test-retest reliability was good across measures; a correlation of 0.70 or greater is generally considered to be acceptable test-retest reliability for clinical trials, and most of the tests were in that range or higher. The one exception was the medication management component of the UCSD Performance-Based Skills Assessment that had relatively low reliability. As mentioned above, this was a secondary measure of the

TABLE 2. Utility as a Repeated Measure in the MATRICS Psychometric and Standardization Study

Co-Primary Measure	Time ₁		Time ₂		Time ₂ -Time ₁ Difference		Number of Scores at Floor/Ceiling	Analysis			Effect Size (Cohen's d)	
	Mean	SD	Mean	SD	Mean	SD		t	df	p		
Maryland Assessment of Social Competence												
Conversational content	3.3	1.0	3.2	1.0	-0.1	0.8	8/9	-2.18	157	0.03	-0.10	
Nonverbal content	3.3	1.1	3.1	1.0	-0.2	0.9	10/25	-3.16	157	0.002	-0.19	
Effectiveness	3.2	1.0	3.1	1.0	-0.1	0.8	11/9	-2.29	165	0.02	-0.10	
UCSD Performance-Based Skills Assessment												
Total score	86.1	11.3	88.6	10.2	2.5	7.2	0/23	4.54	165	0.0001	0.23	
Medication management ^a	7.4	6.6	6.5	6.3	-0.9	6.7	0/48	-1.64	165	0.10	-0.13	
Schizophrenia Cognition Rating Scale												
Patient ^a	4.1	2.1	4.0	2.1	-0.1	1.9	3/38	-0.50	164	0.62	-0.03	
Informant ^a	4.1	1.7	4.3	1.8	0.2	1.2	0/9	1.73	130	0.09	0.10	
Interviewer ^a	4.7	1.5	4.6	1.6	-0.1	0.9	0/2	-1.28	157	0.20	-0.06	
Clinical Global Impression of Cognition in Schizophrenia												
Neurocognitive state (patient) ^a	3.2	0.9	3.1	0.9	-0.1	0.6	0/3	-1.35	164	0.18	-0.08	
Neurocognitive state (caregiver) ^a	3.0	1.0	3.1	0.9	0.1	0.5	0/8	1.34	131	0.18	0.06	
Neurocognitive state (composite) ^a	3.4	0.9	3.3	0.9	-0.1	0.6	0/2	-1.64	158	0.10	-0.08	
Global Assessment of Functioning (composite)	59.2	9.3	59.6	9.2	0.4	4.3	0/0	1.08	160	0.28	0.04	

^a Tests in which lower scores are better.

UCSD Performance-Based Skills Assessment. For the self-report measures, the reliability was slightly higher for the interviewers and informants than for the patients.

Utility as a Repeated Measure

Tests are considered useful for clinical trials if they show a relatively small practice effect or, if they have a practice effect, the effect is not so large that subjects' scores approach the ceiling. To examine a test's utility as a repeated measure, we assessed scores at baseline and 4 weeks later, change scores, the variability of change, and the number of administrations at which the subjects performed at the ceiling or floor, defined as the best or worst score possible on the test (shown in Table 2). Note that in some of the tables, the direction of scoring differs among test indices. Scores for which lower is better are noted. The practice effects were generally small, and the highest was for the UCSD Performance-Based Skills Assessment total score (Cohen's $d=0.23$). Other effect sizes for practice effects with the co-primary measures were small, but some were statistically significant with the large group size. Finally, a few tests showed ceiling effects, including the medication management component of the UCSD Performance-Based Skills Assessment, the patient report on the Schizophrenia Cognition Rating Scale, the nonverbal score from the Maryland Assessment of Social Competence, and the total score from the UCSD Performance-Based Skills Assessment.

Relationship to Cognitive Performance

Because the co-primary measures are intended to serve as face valid indicators of the consequences of underlying

changes in cognition, it is important that they correlate with cognitive performance. We examined the correlations of the co-primary measures with each of the components of the MATRICS Consensus Cognitive Battery (in the data supplement), as well as a composite score from the 10 tests (shown in Table 3). For this purpose, the composite score was calculated by standardizing each of the cognitive tests (mean of 0 and SD of 1) and then summing the standardized scores. The MATRICS Consensus Cognitive Battery scoring program now has a more systematic way to derive a composite score that involves two renorming steps, first for each of the domain scores that have multiple measures and second for the final composite score. We present the simplified composite score in this article because the PASS data were collected and analyzed before the MATRICS Consensus Cognitive Battery scoring program existed, so the data in Table 4 are those that were reviewed by the MATRICS Neurocognition Committee and submitted to NIMH and the FDA. To compare the strength of correlations among measures, we included only subjects with complete data on all four measures (i.e., listwise deletion, $N=156$). The correlation between the UCSD Performance-Based Skills Assessment total score and cognitive performance was significantly higher than for both interview-based summary measures (Schizophrenia Cognition Rating Scale—interviewer and Clinical Global Impression of Cognition in Schizophrenia—neurocognitive state composite) ($t=4.13$, $df=153$, $p=0.001$). The UCSD Performance-Based Skills Assessment correlation with cognitive performance was also higher than the one for the

TABLE 3. Relationship to Cognitive Performance in the MATRICS Psychometric and Standardization Study^a

Co-Primary Measure	Overall Cognition
Maryland Assessment of Social Competence	
Conversational content	0.42
Nonverbal content	0.32
Effectiveness	0.40
UCSD Performance-Based Skills Assessment	
Total score	0.61
Medication ^b	-0.53
Schizophrenia Cognition Rating Scale	
Patient ^b	-0.18
Informant ^b	-0.21
Interviewer ^b	-0.31
Clinical Global Impression of Cognition in Schizophrenia	
Neurocognitive state (patient) ^b	-0.30
Neurocognitive state (caregiver) ^b	-0.24
Neurocognitive state (composite) ^b	-0.31
Global Assessment of Functioning (composite)	0.30

^a Cognitive performance scores that were used: Trail Making Test part A: time to completion; Brief Assessment of Cognition in Schizophrenia symbol coding: total number correct; Hopkins Verbal Learning Test—Revised: total number of words recalled over three learning trials; Wechsler Memory Scale—III: spatial span: sum of raw scores on forward and backward conditions; Letter-Number Span: number of correct trials; Neuropsychological Assessment Battery: mazes: total raw score; Brief Visuospatial Memory Test—Revised: total recall score over three learning trials; Category Fluency (animal naming): total number of animals named in 60 seconds; Mayer-Salovey-Caruso Emotional Intelligence Test: managing emotions: branch score using general consensus scoring; Continuous Performance Test—identical pairs: mean d' value across 2-, 3-, and 4-digit conditions.

^b Tests in which lower scores are better.

Maryland Assessment of Social Competence Effectiveness score ($t=2.50$, $df=153$, $p<0.03$).

Relationship to Self-Reported Functional Outcome

The sites differed substantially in the functional status of the patients (e.g., one site had only a single subject working, and another site consisted largely of patients in a long-term residential program). Because of these cross-site differences in functional status, the correlations between co-primary measures and functional outcome varied from site to site. Given this variability, we present correlations two ways: a correlation for all subjects across sites and also the median correlation across the five sites. Both of these methods tend to underestimate the correlations that would be observed if patients at all sites included a wide range of functional levels. However, the results (Table 4) allow for direct comparisons among the measures. Most of the correlations were modest, and there were no notable differences in the strength of the correlations across co-primary measures.

Tolerability/Practicality

Tolerability refers to the participant's perspective of a measure (i.e., how interesting, pleasant, or burdensome it is to take). Similar to what was done with the cognitive

measures (4), the subjects were asked immediately after each co-primary measure to point to a number on a 7-point Likert scale (1=extremely unpleasant, 7=extremely pleasant) to indicate the degree to which they found the measure pleasant. Practicality refers to the administrator's view of the measure (i.e., how difficult it is to set up, train staff, administer, and score). Practicality was assessed with a 7-point Likert scale for setup, administration, and scoring, as well as a global score. The ratings were made by the administrators of the co-primary measures after completion of data collection for the entire group.

Table 5 presents the data on tolerability and practicality, as well as the mean length of time for administration. For the two functional capacity measures, both tests were considered to be practical and tolerable, with the UCSD Performance-Based Skills Assessment generally rated higher than the Maryland Assessment of Social Competence. We did not record scoring as part of the practicality rating for the Maryland Assessment of Social Competence because that test was scored centrally, not at the local sites. However, a practical consideration is that this test required the additional step of centralized scoring, a process that typically required 20–30 minutes per assessment. For the two interview-based measures, both tests had acceptable ratings, with the Schizophrenia Cognition Rating Scale generally rated higher than the Clinical Global Impression of Cognition in Schizophrenia because it was easier to score and could be administered in roughly half the time. Of course, investigators desiring cognitive domain coverage might nevertheless prefer the Clinical Global Impression of Cognition in Schizophrenia.

Missing Data

Table 6 shows the number of data missing across both assessments, not counting the nine missing assessments for the subjects who did not return for the retest. The number of missing data is quite small for both functional capacity measures, with the UCSD Performance-Based Skills Assessment showing fewer missing data than the Maryland Assessment of Social Competence. For the interview-based measures, the data collection was nearly complete for the patient interviews, but about 14% of the informant interviews (across both assessment periods) were missing. This pattern reflects the expected challenge in identifying and locating suitable informants for the patient participants in clinical trials.

Site Differences

We also evaluated site differences among representative measures from each of the four co-primary measures (effectiveness from the Maryland Assessment of Social Competence, total score from the UCSD Performance-Based Skills Assessment, and the interviewer's ratings from each self-report measure) with a series of one-way analyses of variance (ANOVAs). The UCSD Performance-Based Skills

TABLE 4. Relationships to Functional Status in the MATRICS Psychometric and Standardization Study

Co-Primary Measure	Relationship to Functional Outcome								
	Global ^a		Social ^a		Independent Living ^a		Work ^a		
	Median	r	Median	r	Median	r	Median	r	
Maryland Assessment of Social Competence									
Conversational content	0.28	0.28	0.00	0.00	0.32	0.29	0.16	0.14	
Nonverbal content	0.23	0.20	0.02	0.00	0.29	0.28	0.08	0.11	
Effectiveness	0.27	0.28	0.00	0.03	0.31	0.26	0.15	0.10	
UCSD Performance-Based Skills Assessment									
Total score	0.23	0.29	0.04	0.00	0.19	0.09	0.17	0.19	
Medication management ^b	-0.18	-0.18	-0.01	0.09	-0.07	-0.08	-0.23	-0.26	
Schizophrenia Cognition Rating Scale									
Patient ^b	-0.23	-0.27	-0.28	-0.26	-0.06	-0.03	-0.06	-0.21	
Informant ^b	-0.24	-0.32	-0.16	-0.29	-0.22	-0.09	-0.05	0.05	
Interviewer ^b	-0.34	-0.37	-0.17	-0.08	-0.27	-0.23	-0.16	-0.08	
Clinical Global Impression of Cognition in Schizophrenia									
Neurocognitive state (patient) ^b	-0.31	-0.39	-0.20	-0.19	-0.24	-0.34	-0.10	-0.23	
Neurocognitive state (caregiver) ^b	-0.27	-0.40	-0.05	-0.07	-0.26	-0.37	-0.16	-0.02	
Neurocognitive state (composite) ^b	-0.30	-0.39	-0.12	-0.11	-0.26	-0.31	-0.14	-0.15	
Global Assessment of Functioning (composite)	0.39	0.41	0.20	0.21	0.26	0.19	0.22	0.18	

^a First value combines data across sites; second value is the median r of the five sites.

^b Tests in which lower scores are better.

TABLE 5. Practicality and Tolerability in the MATRICS Psychometric and Standardization Study

Co-Primary Measure	Practicality ^a										Administration Time	
	Global		Setup		Administration		Scoring		Tolerability ^b		Mean	SD
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
Maryland Assessment of Social Competence	4.7	1.2	5.8	0.8	4.3	1.4	—	—	3.8	2.2	23.8	4.9
UCSD Performance-Based Skills Assessment	5.8	0.8	5.2	1.2	5.8	1.2	6.2	1.0	6.2	1.0	25.7	8.7
Schizophrenia Cognition Rating Scale	6.0	0.9	7.0	0.0	6.2	0.8	5.8	0.8	5.7	1.0	13.3	3.8
Clinical Global Impression of Cognition in Schizophrenia	3.5	0.8	6.0	1.1	3.5	1.0	3.0	0.6	3.5	0.8	25.8	3.8

^a Possible range=1 (very hard) to 7 (very easy).

^b Possible range=1 (extremely unpleasant) to 7 (extremely pleasant).

Assessment did not show significant differences across sites, but the other three measures did (Maryland Assessment of Social Competence effectiveness: $F=3.39$, $df=4$, 161, $p<0.03$; Schizophrenia Cognition Rating Scale—interviewer: $F=6.67$, $df=4$, 165, $p<0.001$; Clinical Global Impression of Cognition in Schizophrenia composite: $F=20.73$, $df=4$, 165, $p<0.001$). The differences across sites on these co-primary measures may be partially a result of the fact that the sites differed in terms of the average functional level of the patients. Hence, we view the differences as a reflection of actual group differences as opposed to a failure of cross-site reliability of administration.

Discussion

This article evaluated four potential co-primary measures that may be used in clinical trials of cognition-enhancing drugs for schizophrenia. Early discussions with the FDA indicated that approval of a cognition-enhancing drug for schizophrenia would require changes on a consensus measure of cognitive performance, as well as on a functionally meaningful co-primary measure. We considered four

different potential co-primary measures, two measures each from two distinctive approaches: measures of functional capacity and interview-based measures of cognition.

Several summary points can be drawn from these data. First, all four measures had acceptable test-retest reliability. An added component of the UCSD Performance-Based Skills Assessment (medication management) had lower reliability, perhaps because scores on this component tended to be at or near the ceiling. Second, most tests had adequate range. A practice effect was noted for the UCSD Performance-Based Skills Assessment total score. The measures with the most prominent ceiling effects included the medication management component of the UCSD Performance-Based Skills Assessment (which is separate from the main test) and the Schizophrenia Cognition Rating Scale—patient rating (which is not the primary rating for this interview). Third, the relationships to cognitive performance were notably higher for functional capacity measures (in particular the UCSD Performance-Based Skills Assessment) than for the interview-based measures of cognition. This pattern is consistent with findings that interview-based assessments of cognition

TABLE 6. Missing Data in the MATRICS Psychometric and Standardization Study

Co-Primary Measure	Tally of Missing Data ^a
Maryland Assessment of Social Competence	18
UCSD Performance-Based Skills Assessment	5
Schizophrenia Cognition Rating Scale—patient	3
Schizophrenia Cognition Rating Scale—informant	49
Clinical Global Impression of Cognition in Schizophrenia—patient	5
Clinical Global Impression of Cognition in Schizophrenia—caregiver	49

^a Missing data across both assessments.

generally do not correlate well with objective cognitive performance (15). Also, the UCSD Performance-Based Skills Assessment, the Maryland Assessment of Social Competence, and the MATRICS Consensus Cognitive Battery can all be considered performance measures, and therefore they share some method variance. Fourth, all measures had modest relationships to community functional status that were somewhat lower than expected. Fifth, missing data were more frequently observed for the interview-based measures because of the difficulty in contacting individuals who could serve as informants.

The blind across measures was not absolute. The raters who administered the functional capacity measures also administered the cognitive performance tests (although the scoring for the Maryland Assessment of Social Competence was done centrally). The testers did not have access to any information on functional status or the interview-based measures of cognition. The staff members who conducted the interview-based measures of cognition did not have access to the functional status interviews, but they were not fully blind to functional status because these measures include questions about the degree to which cognitive impairment interferes with activities of daily living. Another limitation of these data are that they were collected in the absence of cognitive change. One would ideally want to know the sensitivity of these measures to changes in cognition as opposed to measuring when the cognition was presumably stable. Questions about the sensitivity of co-primary measures to change in cognitive performance will await identification of a potent cognition-enhancing drug or perhaps could be examined in the context of cognitive remediation programs.

The relationships between the co-primary measures and functional status were somewhat lower than expected and lower than reported in other studies (10, 16). The ratings of community functioning were based entirely on subject self-reports (not observation and not informants), which may have limited the validity of the ratings. In addition, the correlations may have been limited by substantial differences in functional status across the five sites and considerable variability in the size of the correlations across sites that may be attributable to instances of within-site com-

munity functioning homogeneity. At some sites, the range of community functioning was restricted by treatment setting constraints (e.g., residential treatment) that were not simply due to cognitive or functional capacity levels. A restricted range of within-site variance on the community functional status measure would be expected to reduce the magnitude of the correlations with co-primary measures. In this situation, it was difficult to obtain a representative indication of the associations; however, the values do provide a reasonable basis for comparison of the four measures, all of which performed comparably.

The MATRICS Neurocognition Committee reviewed the data presented in this article and concluded that because all of the potential co-primary measures performed reasonably well, it was difficult to make a narrow recommendation to NIMH and the FDA. A clearly stated objective for the MATRICS Initiative was to select a single consensus cognitive performance battery (the MATRICS Consensus Cognitive Battery) for use in clinical trials of cognition-enhancing drugs for schizophrenia. In contrast, the Neurocognition Committee was not asked to recommend any particular co-primary measure, and the committee did not do so. The Neurocognition Committee noted that if one needed to choose a co-primary measure for a clinical trial at this time, the UCSD Performance-Based Skills Assessment (with modifications to address the ceiling effect) had the advantage of a strong association with cognitive performance. A revised version of the UCSD Performance-Based Skills Assessment with adjusted difficulty level is currently available. The two interview-based measures were similar in structure, and their summary scores were highly correlated ($r=0.68$; in comparison, the correlation for the two functional capacity measures was 0.30). If it was desirable to include an additional co-primary measure that reflects the interview-based approach, the Schizophrenia Cognition Rating Scale has the advantage of better tolerability and practicality, although it lacks domain ratings.

The evaluation and selection of co-primary measures raises questions about validity. The key concern of the FDA is demonstrating that a new cognition-enhancing drug affects cognitive performance and meaningful functional abilities. At present, evidence of content validity of co-primary measures may be more available than evidence of construct or external validity. That is, the content of these measures can be evaluated for its functional meaningfulness from the perspectives of patients, clinicians, and families. Further establishment of construct or external validity, including relationships to cognitive performance and community functioning, will take time. A related validity issue is whether co-primary measures need to be sufficiently distinct (i.e., involve different constructs) from cognitive performance measures. For example, some functional capacity measures resemble neurocognitive tests in mode of administration. However, at this point, distinctiveness of cognitive performance and co-primary measures does not appear to be a concern of

the FDA. A co-primary measure could in theory be a modified version of a cognitive performance test (if it has content of clear functional relevance) or a modified version of a community functioning scale (if sufficiently sensitive to change within a clinical trial). At this stage of development, associations between co-primary measures and cognitive performance may be more critical than those with everyday functional outcome, given that improvements in functional capacity may still require additional rehabilitation in specific daily living or work skills to be translated into community functioning.

Representatives of the FDA indicated to the MATRICS Neurocognition Committee that any of the four co-primary measures would be acceptable at this point for use in clinical trials. However, given the early stage of method development in this area, the Neurocognition Committee expects that stronger measures using these approaches, or using entirely different approaches to assessment, might be developed. This is a fertile area and one in which developments are occurring at a rapid pace. The importance of functional assessment for serious mental disorders and its role in drug evaluation has been reflected in several recent developments. For example, a new program at NIMH, Functional Assessment in Mental Disorders, has recently been created. Also, a new academic/industry consortium (MATRICS-Co-Primary and Translation) has been formed to facilitate the evaluation of potential co-primary measures for use in clinical trials. Finally, NIMH recently sponsored a consensus meeting on functional assessment for psychiatric disorders (21). Hence, the results presented in the current article should be viewed as a good starting point for evaluation, comparison, and discussion of potential functionally meaningful co-primary measures, with expectations that future developments will lead to improved tools in this domain.

Received Jan. 13, 2007; revisions received July 6 and Aug. 12, 2007; accepted Aug. 16, 2007 (doi:10.1176/appi.ajp.2007.07010089). From the Semel Institute for Neuroscience and Human Behavior and the Department of Psychology, Geffen School of Medicine, UCLA; the VA Greater Los Angeles Healthcare System, Los Angeles; the University of Kansas School of Medicine, Wichita, Kan.; NIMH, Bethesda, Md.; the Maryland Psychiatric Research Center, University of Maryland School of Medicine, Baltimore; the Department of Psychiatry and Behavioral Sciences, Duke University, Durham, N.C.; the Public Psychiatry Division, Department of Psychiatry, Beth Israel Deaconess Medical Center, Massachusetts Mental Health Center, Boston; and the Department of Psychiatry, Harvard Medical School, Boston. Address correspondence and reprint requests to Dr. Green, Semel Institute for Neuroscience and Human Behavior, UCLA, 300 Medical Plaza, Rm. 2263, Los Angeles, CA 90095-6968; mgreen@ucla.edu (e-mail).

Dr. Keefe has received funding from Abbot, AstraZeneca, Dainippon Sumitomo, Forest, Gabriel, GlaxoSmithKline, Memory, Otsuka, Pfizer, Repligen, Saegis, Bristol-Myers Squibb, Cephalon, Eli Lilly, Johnson & Johnson, Lundbeck, Merck, Orexigen, Sanofi-Aventis, Xenoport. He receives royalties from the Brief Assessment of Cognition and the MATRICS Battery. He has received grants from Organon, NIMH, Pfizer, Eli Lilly, and AstraZeneca. Dr. Green has consulted for Eli Lilly, Otsuka, Lundbeck, Bristol-Myers Squibb, Abott, Sanofi-Aventis, Memory, Solvay, and Astellas. Dr. Marder has received funds from Bristol-Myers Squibb, Otsuka, Solvay, Pfizer, Merck, Roche, and Wy-

eth. Dr. Gold receives royalties from the Brief Assessment of Cognition. Dr. Neuchterlein has received a grant from Janssen.

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

The authors thank Dr. Jim Mintz and the UCLA Semel Institute Biostatistical Unit for developing the data entry systems and Bi-Hong Deng for her help in data management and preparation of the tables. The following staff members at the UCLA-VA site coordinated the training and quality assurance procedures: Karen Cornelius, Psy.D., Kimmy Kee, Ph.D., Mark McGee, Ayala Ofek, and Mark Sergi, Ph.D. Training and quality assurance for the co-primary measures was provided by Alan Bellack, Ph.D., Robert Bilder, Ph.D., Sherry Goldman, M.A., Thomas Patterson, Ph.D., Margaret Poe, Shannon Thomas-Lohrman, M.S., and Joseph Ventura, Ph.D.

Dr. Fenton died in September 2006.

Acknowledgment

Other members of the MATRICS Outcomes Committee included Jean Addington, Robert Heinssen, Richard Mohs, Thomas Patterson, and Dawn Velligan.

References

1. 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
2. Green MF, Nuechterlein KH, Gold JM, Barch DM, Cohen J, Essock S, Fenton WS, Frese F, Goldberg TE, Heaton RK, Keefe RSE, Kern RS, Kraemer H, Stover E, Weinberger DR, Zalcman S, Marder SR: Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. *Biol Psychiatry* 2004; 56:301–307
3. Buchanan RW, Davis M, Goff D, Green MF, Keefe RSE, Leon AC, Nuechterlein KH, Laughren T, Levin R, Stover E, Fenton W, Marder SR: A summary of the FDA-NIMH-MATRICS workshop on clinical trial design for neurocognitive drugs for schizophrenia. *Schizophr Bull* 2005; 31:5–19
4. Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch D, Cohen J, Essock S, Fenton WS, Frese FJ, Gold JM, Goldberg T, Heaton R, 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 January 2, 2008; doi: 10.1176/appi.ajp.2007.07010042)
5. Kern RS, Nuechterlein KH, Green MF, Baade LE, Fenton WS, Gold JM, Keefe RSE, Mesholam-Gately R, Mintz J, Seidman LJ, Stover E, Marder SR: The MATRICS Consensus Cognitive Battery: part 2, conorming and standardization. *Am J Psychiatry* (published online January 2, 2008; doi: 10.1176/appi.ajp.2007.07010043)
6. Nuechterlein KH, Green MF: MATRICS Consensus Cognitive Battery. Los Angeles, MATRICS Assessment, 2006
7. Brekke JS, Kay DD, Kee KS, Green MF: Biosocial pathways to functional outcome in schizophrenia. *Schizophr Res* 2005; 80: 213–225
8. Vauth R, Rusch N, Wirtz M, Corrigan PW: Does social cognition influence the relation between neurocognitive deficits and vocational functioning in schizophrenia? *Psychiatry Res* 2004; 128:155–165
9. Sergi MJ, Rassovsky Y, Nuechterlein KH, Green MF: Social perception as a mediator of the influence of early visual process-

- ing on functional status in schizophrenia. *Am J Psychiatry* 2006; 163:448–454
10. Bowie CR, Reichenberg A, Patterson TL, Heaton RK, Harvey PD: Determinants of real-world functional performance in schizophrenia subjects: correlations with cognition, functional capacity, and symptoms. *Am J Psychiatry* 2006; 163:418–425
 11. Brekke JS, Long JD, Nesbitt N, Sobel E: The impact of service characteristics from community support programs for persons with schizophrenia: a growth curve analysis. *J Consult Clin Psychol* 1997; 65:464–475
 12. 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
 13. Bellack AS, Sayers M, Mueser K, Bennett M: Evaluation of social problem solving in schizophrenia. *J Abnorm Psychol* 1994; 103:371–378
 14. Dunning D, Heath C, Suls JM: Flawed self-assessment: implications for health, education, and the workplace. *Psychol Science in the Public Interest* 2004; 5:69–106
 15. Mortiz S, Ferahli S, Naber D: Memory and attention performance in psychiatric patients: lack of correspondence between clinician-rated and patient-rated functioning with neuropsychological test results. *J Int Neuropsychol Soc* 2004; 10: 623–633
 16. Keefe RS, 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
 17. Ventura J, Cienfuegos A, Boxer O, Bilder RM: Clinical Global Impression of Cognition in Schizophrenia (CGI-CogS): reliability and validity of a co-primary measure of cognition. *Schizophr Res* (in press)
 18. 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
 19. 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
 20. Weissman M, Paykel E: *The Depressed Woman: A Study of Social Relationships*. Chicago, University of Chicago Press, 1974
 21. Bellack AS, Green MF, Cook JA, Fenton W, Harvey PD, Heaton RK, Laughren T, Leon AC, Mayo D, Patrick DL, Patterson TL, Rose A, Stover E, Wykes T: Assessment of community functioning in people with schizophrenia and other severe mental illnesses: a white paper based on an NIMH-sponsored workshop. *Schizophr Bull* 2007; 33:805–822