The Reliability and Validity of the MATRICS Functional Assessment Battery

Dawn I. Velligan[*,1], Megan Fredrick[1], Jim Mintz[1], Xueying Li[1], Maureen Rubin[2], Sanjay Dube[3], Smita N. Deshpande[4], Jitendra K. Trivedi[5], Shiv Gautam[6], Ajit Avasthi[7], Robert S. Kern[8], and Stephen R. Marder[9]

[1]Department of Psychiatry, University of Texas Health Science Center at San Antonio, San Antonio, TX; [2]Department of Social Work, University of Texas at San Antonio, San Antonio, TX; [3]Eli Lilly and Company, Indianapolis, IN; [4]Postgraduate Institute of Medical Education & Research, Dr. Ram Manohar Lohia Hospital, New Delhi, India; [5]CSM Medical University, Lucknow, India; [6]Gautam Institute of Behavioral Sciences and Alternative Medicine, NIMS Medical College and University, Jaipur, India; [7]Postgraduate Institute of Medical Education & Research, Chandigarh, India; [8]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA; [9]Semel Institute for Neuroscience, University of California, Los Angeles and VA Desert Pacific Mental Illness Research, Education and Clinical Center, Los Angeles, CA

*To whom correspondence should be addressed; Department of Psychiatry, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, US; tel: 210-567-5508, fax: 210-567-129, e-mail: velligand@uthscsa.edu

Objectives: The Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative was designed to encourage the development of cognitive enhancing agents for schizophrenia. For a medication to receive this indication, regulatory agencies require evidence of improvement in both cognition and functional outcome. Functional capacity measures typically used in clinical trials as intermediate measures of functional outcome must be adapted to fit different cultural contexts for use internationally. We examined the psychometric properties of the MATRICS Functional Assessment Battery (MFAB), comprised of 2 subtests from the UCSD Performance-based Skills Assessment (UPSA) and one from the Test of Adaptive Behavior in Schizophrenia (TABS) that were rated by experts in a previous study to be the most appropriate functional capacity assessments across different cultural contexts. *Method:* Four sites in India administered the MFAB, a brief version of the UPSA, the MATRICS Cognitive Consensus Battery, measures of symptomatology, and a measure of global functional outcome to 141 individuals with schizophrenia at a baseline assessment and at 4 weeks later. *Results:* Test-retest reliability based on the intraclass correlation coefficient was significantly better for the UCSD Performance-Based Skills Assessment-Brief (UPSA-B). Pearson correlation coefficients over time were not significantly different for the 2 measures. Only the MFAB was significantly correlated with functional outcome as measured by the Specific Levels of Functioning Scale. *Conclusions:* The psychometric properties of the MFAB and UPSA-B were similar. The MATRICS scientific board chose to translate the MFAB into multiple languages for potential use in studies of novel medications seeking an indication for improving cognition in schizophrenia.

Key words: schizophrenia/co-primary measures/ intermediate measures/international clinical trials/ functional capacity measures/cognitive impairments

Introduction

The Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative was designed to encourage the development of cognitive enhancing pharmaceutical agents for schizophrenia by developing a process by which a medication could receive an indication for the treatment of cognitive dysfunction in schizophrenia.[1–5] This initiative was collaboration among academicians, industry partners, and government agencies and resulted in recommendations for study design and the development of a consensus cognitive battery—the MATRICS Consensus Cognitive Battery (MCCB)—to assess cognition in studies of novel compounds seeking this indication.[4,5] Representatives of the US Food and Drug Administration (FDA) indicated that improvement in performance on neuropsychological tests was not sufficient to establish an indication for improving cognition in schizophrenia.[3] The FDA indicated that a compound would also need to demonstrate that it improved a co-primary measure of functional outcome that had more face validity for everyday functioning than cognitive testing.[3,6]

As part of the MATRICS initiative, the Validation of Intermediate Measures (VIM) study assessed the reliability, validity, and utility of a number of intermediate measures assessing functional outcome in schizophrenia.[7] Because longer term functional outcomes such as employment, or changes in marital status are not likely to be improved during the course of a typical clinical trial,

the study focused on intermediate measures of functional capacity or everyday functioning, which are thought to be more amenable to change over this time period.[6] Findings indicated that the UCSD Performance-Based Skills Assessment (UPSA)[8] and brief versions of the UPSA and Test of Adaptive Behavior in Schizophrenia (TABS)[9] were the instruments with the most favorable psychometric properties in a US sample. However, many efficacy studies of novel compounds are now conducted as multisite international trials.[10]

In the cross-cultural adaptability of intermediate measures (CIM) study, we examined which of the intermediate measures from the VIM study would be rated as most appropriate for use across cultures by expert investigators conducting clinical trials in 8 different countries.[10] We obtained opinions regarding the overall adaptability of each intermediate measure and its applicability across genders, socioeconomic strata, ethnicity, and geographic region (rural vs urban) for patients typically seen at international sites. Two subtests from the UPSA and one from the TABS were rated as most appropriate for adaptation across multiple countries and were combined into the MATRICS Functional Assessment Battery (MFAB). The present study sought to evaluate the psychometric properties of the MFAB in India. Although experts in Mexico and China also identified challenges in adaptation, India was the country with the most striking challenges according to the results of the CIM study.[10] We reasoned that if the MFAB demonstrated acceptable psychometric properties in the country with the most difficult issues in adaptation, then this would be a good indication that the measure could be adapted for use in international clinical trials as a co-primary measure of functional outcome and could be used successfully in less westernized, less industrialized countries in general.

We examined internal consistency, test-retest reliability, and criterion validity of the MFAB compared to a version of the UCSD Performance-Based Skills Assessment-Brief (UPSA-B) currently being utilized in studies in India. The purpose of the trial was to select for translation a functional capacity measure that would be culturally acceptable for use in both developed and less-developed countries to be used as a co-primary measure for studies of medications seeking a clinical indication for cognitive enhancement in schizophrenia. Our primary criteria were reliability over time and correlation with cognition and global functional outcome. Internal consistency is less important in real-world applications where the goal is to predict criterion variables. In this situation, it is often more appropriate to use composite measures of different skills.[11]

The UPSA-B was selected as a comparator measure because it is currently in use in studies in India. The UPSA-B demonstrated adequate psychometric properties in a US-only study of reliability and validity. None of the tests in the brief version of the UPSA overlap with the

tests from the full UPSA, some of which were selected for inclusion in the FAB.[7] Subtests on these functional capacity measures do not assess discrete functions like some neuropsychological tests do, but rather are designed to apply integrated cognitive abilities to everyday scenarios. We anticipated at least moderate correlations between the MFAB and the UPSA-B despite the fact that none of the subtests overlapped.

Method

Study Design

One hundred and sixty Hindi-speaking participants were recruited at 4 sites in India from April to December 2012. After signing informed consent, participants were assessed at baseline and 4 weeks postbaseline on the FAB and measures of cognitive functioning, symptoms, and functional outcome. This research was approved by the Institutional Review Board of the University of Texas Health Science Center, which was the coordinating site for the study, as well as by the India Council of Medical Research.

Sites

Sites in India were recommended for participation by industry members of the MATRICS scientific board composed of members from academia, industry, and the National Institute of Mental Health and were nominated based upon their experience in conducting clinical trials in schizophrenia and administering cognitive and functional assessments. Clinical investigators at 6 sites were trained in the administration and scoring of instruments and rating scales at a central meeting in India. Four sites located in New Delhi, Jaipur, Chandigarh, and Lucknow recruited participants for the study. Online quality management was conducted throughout the study to ensure accurate scoring and administration of testing.

Participants

Inclusion criteria were identical to those in the VIM study with the exception of language[7]: (1) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, diagnosis of schizophrenia, outpatient status, age 18–60, ability to understand and read Hindi sufficiently to comprehend testing procedures, ability to comprehend the consent form, had not received the performance-based intermediate measures in this study, the MCCB or similar cognitive assessment within 6 months of study entry, and clinical stability as evidenced by no significant psychotropic medication changes in past 2 months and none anticipated for the next month, (2) evidence of stable symptomatology ≥3 months, (3) Positive and Negative Syndrome Scale (PANSS) score no more than 4 (moderate) on P1 delusions, P2 conceptual disorganization, P3 hallucinatory behavior, P5 grandiosity, P6 Suspiciousness,

and G8 unusual thought content, (4) PANSS score no more than 15 total on Negative Syndrome subscale, and (5) mood symptoms, if present, had been stable for at least 3 months.

Patients were excluded if they had alcohol or substance dependence in the past 6 months, alcohol or substance abuse in the past 3 months, clinically significant neurological disease, head injury (eg, loss of consciousness over 1 h), a current medical condition that would interfere with valid assessment, dystonia or parkinsonism that would affect validity of assessment, or were taking any of the following medications: clozapine, potentially procognitive medications, antidementia medications, amphetamine, lithium, monoamine oxidase inhibitors, tricyclic antidepressants. No benzodiazepines, sedatives, or anticholinergic medications were administered within 12 hours of assessment. After complete description of the study to the subjects, written informed consent was obtained.

Assessments

Functional Capacity Measures. The MFAB[12] is comprised of 3 test areas. Two test areas Comprehension and Household Management were chosen from the UPSA[8] and 1 test area Work and Productivity was chosen from the TABS.[9] The MFAB assesses a variety of adaptive skills needed for daily functioning. MFAB scores for each area are calculated as percent correct (ie, points for each assessment were added together and divided by the total possible points for the area and then multiplied by 100) or are transformed to yield scores ranging from 0 to 100. A mean MFAB score is calculated. Higher scores indicated better adaptive functioning.

UCSD Performance-Based Skills Assessment-Brief. The UPSA-B[8] was designed to assess the abilities of individuals to perform everyday tasks that are considered necessary for independent functioning in the community. The UPSA-B assesses 2 skill areas that are considered essential to functioning in the community: Finance and Social/Communications. The raw scores from each of the subscales are transformed to yield comparable scores (0–20) for each scale. Scores are combined into a summary score with higher scores reflecting better performance. None of the UPSA-B subtests overlap with subtests on the MFAB.

Functional Outcome. The Specific Levels of Functioning Scale (SLOF)[13] assesses interpersonal relationships and a number of community living skills on a series of 5-point scales based upon an interview with the patient and caregiver. Higher scores indicate better functioning. A mean score from the SLOF reflects a global level of community functioning.

Cognition. The MCCB assesses 7 separable cognitive domains including speed of processing, attention/vigilance,

working memory, verbal learning, visual learning, reasoning and problem solving, and social cognition.[4] A global cognitive summary score from the MATRICS battery is used as the primary measure of cognitive functioning.

Symptomatology. Symptoms were rated on the PANSS.[14] Positive and negative symptom factors were created based upon the work of Marder et al.[15] Higher scores reflect greater psychopathology.

Statistical Analysis

Analyses of reliability were based on $N = 141$, who completed both the MFAB and UPSA at both assessments. All the statistics were based on the same sample, so comparisons of MFAB and UPSA required formulas for dependent statistics. Internal consistency reliability was evaluated using Cronbach's alpha and compared between measures using the test by Feldt[16] for dependent alphas. Test-retest reliability was assessed using Pearson correlation coefficients and compared between measures using a modification of the Pearson-Filon method.[17,18] Intraclass correlations were calculated as the proportion of total variance accounted for by between-subjects variance without adjusting for mean differences between assessment occasions and compared between measures using the method for comparing dependent ICCs described by Donner and Zou.[19] Validity coefficients were Pearson correlation coefficients, and compared using Hotelling's t test for correlations that share a variable described by McNemar.[20(p140)]

Results

Enrollment

A total of 175 participants were screened and consented. Of these, 148 had complete data at baseline 141 were reassessed at week 4. Each site recruited between 20% ($n = 29$) and 30% ($n = 44$) of the total participants for the study. The baseline demographic and clinical characteristics of the participants appear in table 1. Scores on the

Table 1. Baseline Characteristics of Participants

% Male (n)	67% (99)
Mean age	32.47 (9.11)
Education	11.06 (3.222)
PANSS positive	1.58 (0.51)
PANSS negative	1.51 (0.48)
MCCB mean score	0.002 (0.63)
MFAB total score	60.57 (12.53)
UPSA-B total score	67.58 (16.89)

Note: PANSS, Positive and Negative Syndrome Scale; MCCB, MATRICS Consensus Cognitive Battery; MFAB, MATRICS Functional Assessment Battery; UPSA-B, UCSD Performance-Based Skills Assessment-Brief.

PANSS indicate that the sample was clinically stable over the 4-week period.

As in the VIM study, priority was given to psychometric features that are most relevant to use of co-primary measures in clinical trials: test-retest reliability and criterion validity (relationships among co-primary measures and cognition, and interview-based measures of functional outcome).

Internal Consistency and Reliability

Internal consistency computed using Cronbach's alpha was 0.58 for the MFAB and 0.77 for the UPSA. The difference between these 2 coefficients was statistically significant ($t = 4.39$, $df = 146$, $P < .00003$).[16] With respect to test-retest reliability, both Pearson correlation coefficients and intraclass correlation coefficients (ICC) were calculated between the initial scores and scores obtained at week 4 (examining consistency with subjects and times of testing considered random effects).[21] Pearson correlation coefficients assess the degree to which the relative rank for the global scores remains similar over time, whereas ICCs assess whether observed scores remain similar over time.[22] The Pearson correlation coefficients over time for the MFAB and UPSA-B were 0.62 ($N = 141$, $P < .0001$) and 0.72 ($N = 141$, $P < .0001$), respectively. These correlations were not statistically significantly different, $z = 1.78$, $P < .075$. However, only the UPSA-B had a test-retest reliability over the 0.70, which is a typically recommended cut off for measures of this type. ICCs over time were 0.60 and 0.72 for the MFAB and UPSA-B, respectively. These ICCs were significantly different ($z = 2.05$, $P = 0.041$).

Validity

Relationships among functional capacity measures, the cognition global score, and interview-based measures of functional outcome appear in table 2. Results indicate that both the MFAB and UPSA were moderately to strongly correlated with the MCCB, and neither was correlated with measures of symptomatology. Only the MFAB was significantly correlated with global functional outcome on the SLOF ($r = .28$; $P < .0006$). There

was a trend ($t = 1.89$, $df = 138$, $P < .06$) for the correlation between the MFAB and the SLOF to be significantly higher than that between the UPSA-B and SLOF ($r = .13$; $P > .11$).

Discussion

In this 4-site assessment-only trial in India, we examined the test-retest reliability and validity of the MFAB in comparison with the UPSA-B for the assessment of functional capacity in clinical trials. Results indicated that the UPSA-B had better internal consistency than the MFAB. This is not surprising since the UPSA-B subtests come from a single test where as the MFAB is a battery comprised of subtests from 2 different tests of functional capacity. Although the internal consistency of neither test was strong, the importance of reliability over time and validity were emphasized in test selection. The test-retest reliability of the MFAB and the UPSA-B were comparable, with the UPSA-B having significantly better reliability on 1 of 2 measures. Both the MFAB and the UPSA-B were moderately correlated with the MCCB demonstrating convergent validity. The magnitude of these correlations is consistent with published data on both measures.[10] Because the correlation between the MFAB or UPSA and the MATRICS battery (or UPSA-B) is constrained by the reliabilities of the instruments, the association may account for more than 80% of the reliable variance in these 2 instruments. This result calls into question, from a statistical perspective, the need for intermediate measures of functional capacity. Functional capacity measures may simply be more face valid attempts to assess underlying cognitive abilities. Of the functional capacity measures, only the MFAB was significantly correlated with the SLOF (the measure of global functional outcome utilized in the current study) providing further evidence of convergent validity. However, in this trial, there were stronger correlations between the MATRICS battery and the SLOF and between the PANSS and the SLOF than there were between the SLOF and either intermediate assessment of functional capacity. The correlation between the PANSS and the SLOF may be due in part to method variance as both are interview-based

Table 2. Correlation Among Functional Capacity Measures, Cognition, Symptoms, and Interviewer Rated Measures of Functional Outcome

	UPSA	MCCB	SLOF	PANSS Positive	PANSS Negative
MFAB	0.53**	0.64**	0.28*	−0.01	0.01
UPSA		0.58**	0.13	0.13	0.13
MCCB			0.50**	−.05	−0.11
SLOF				−47**	−0.42**

Note: Abbreviations are explained in the first footnote to table 1. SLOF, Specific Levels of Functioning Scale.
*$P < .001$; **$P < .0001$.

measures and in part to the impact of symptoms of the types of functioning assessed in the SLOF. The correlation between the MATRICS battery and the SLOF is additional evidence that impairments in neurocognition have functional consequences.[23] Although the FDA has indicated that functioning in trials designed to assess cognitive outcomes must be measured with an instrument that has greater face validity for assessing community functioning than neuropsychological tests, these results raise important questions about this decision. The lack of a relationship among the functional capacity measures and positive symptoms is evidence of discriminant validity. Overall results suggest fairly comparable psychometric properties for the UPSA-B and MFAB.

Strengths of the study include that it was conducted in India, one of the ex-US countries that rated cultural adaptability to be the most challenging for tests of functional capacity. Weaknesses of the study included that sites were from only 1 ex-US country. However, the feedback from experts in Mexico and China about the types of problems with the cultural adaptation of these functional capacity tests was very similar to that obtained from experts in India. Therefore, the performance of measures in India is likely to be comparable with their performance in other countries that are dissimilar culturally to the United States. Finally, because this was not a treatment trial, the sensitivity to change of these functional capacity measures was not assessed and will be important to examine in future trials.

Recommendations

The psychometric properties of the MFAB and UPSA-B were similar. A previous study indicated that the tests included in the MFAB battery were rated by experts as more culturally adaptable and more similar to daily activities in multiple countries than those in the UPSA-B.[10] Therefore, based upon all the available information, the MATRICS scientific board chose the MFAB as the measure of functional capacity to translate into multiple languages for use in clinical trials as a co-primary measure. Other measures of functional outcome such as global measures are also viewed as acceptable by regulatory agencies, so the MFAB is not considered to be the only measure that can be used for studies seeking an indication for improvements in cognitive functioning.[3] However, the results of our studies support that the MFAB is a culturally adaptable measure of functional capacity with adequate psychometric properties and relationships with cognitive performance and functional outcomes. Future research examining functional outcome measures based upon behavioral sampling using electronic devices may be important to pursue, as well as virtual reality measures of functional outcome.[24,25]

References

1. Green MF, Nuechterlein KH, Gold JM, et al. Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. *Biol Psychiatry*. 2004;56:301–307.
2. Nuechterlein KH, Robbins TW, Einat H. Distinguishing separable domains of cognition in human and animal studies: what separations are optimal for targeting interventions? A summary of recommendations from breakout group 2 at the measurement and treatment research to improve cognition in schizophrenia new approaches conference. *Schizophr Bull*. 2005;31:870–874.
3. Buchanan RW, Davis M, Goff D, et al. 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, et al. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am J Psychiatry*. 2008;165:203–213.
5. Kern RS, Nuechterlein KH, Green MF, et al. The MATRICS Consensus Cognitive Battery, part 2: co-norming and standardization. *Am J Psychiatry*. 2008;165:214–220.
6. Green MF, Nuechterlein KH, Kern RS, et al. Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS Psychometric and Standardization Study. *Am J Psychiatry*. 2008;165:221–228.
7. Green MF, Schooler NR, Kern RS, et al. Evaluation of functionally meaningful measures for clinical trials of cognition enhancement in schizophrenia. *Am J Psychiatry*. 2011;168:400–407.
8. 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.
9. Velligan DI, Diamond P, Glahn DC, et al. The reliability and validity of the Test of Adaptive Behavior in Schizophrenia (TABS). *Psychiatry Res*. 2007;151:55–66.
10. Velligan DI, Rubin M, Fredrick MM, et al. The cultural adaptability of intermediate measures of functional outcome in schizophrenia. *Schizophr Bull*. 2012;38:630–641.
11. Nunnally JC. *Psychometric Theory*. New York: McGraw Hill; 1967.
12. Patterson TL, Velligan DI. *MATRICS Functional Assessment Battery*. Los Angels, CA: MATRICS Assessment Inc.; 2012.

13. Schneider LC, Struening EL. SLOF: a behavioral rating scale for assessing the mentally ill. *Soc Work Res Abstr*. 1983;19:9–21.

14. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261–276.

15. Marder SR, Davis JM, Chouinard G. The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: combined results of the North American trials. *J Clin Psychiatry*. 1997;58:538–546.

16. Feldt LS. A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika*. 1980;45:99–105.

17. Raghunathan TE, Rosenthal R, Rubin DB. Comparing correlated but nonoverlapping correlations. *Psychol Methods*. 1996;1:178–183.

18. Kenny DA. *Correlation and Causality*. New York: John Wiley & Sons; 1979.

19. Donner A, Zou G. Testing the equality of dependent intra-class correlation coefficients. *J R Stat Soc Series D (The Statistician)*. 2002;51:367–379.

20. McNemar Q. *Psychological Statistics*. 3rd ed. New York: Wiley; 1962.

21. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.

22. Wilk CM, Gold JM, Bartko JJ, et al. Test-retest stability of the Repeatable Battery for the Assessment of Neuropsychological Status in schizophrenia. *Am J Psychiatry*. 2002;159:838–844.

23. Green MF. What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry*. 1996;153:321–330.

24. Harvey PD, Se Keefe R. Technology, society, and mental illness: challenges and opportunities for assessment and treatment. *Innov Clin Neurosci*. 2012;9:47–50.

25. Granholm E, Loh C, Swendsen J. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. *Schizophr Bull*. 2008;34:507–514.