Published in final edited form as:

Schizophr Res. 2010 June ; 119(1-3): 246–252. doi:10.1016/j.schres.2010.03.021.

Psychometric Properties of Performance-based Measurements of Functional Capacity: Test-Retest Reliability, Practice Effects, and Potential Sensitivity to Change

Feea R. Leifker, MPH^a, Thomas L. Patterson, PhD^b, Christopher R. Bowie, PhD^c, Brent T. Mausbach, PhD^b, and Philip D. Harvey, PhD^d

^aDepartment of Psychology, The Pennsylvania State University, State College, PA, USA

^bDepartment of Psychiatry, University of California, San Diego, San Diego, CA, USA

^cDepartment of Psychology, Queens University, Kingston, Ontario, Canada

^dDepartment of Psychiatry, Emory University School of Medicine, Atlanta, GA, USA

Abstract

Performance-based measures of the ability to perform social and everyday living skills are being more widely used to assess functional capacity in people with serious mental illnesses such as schizophrenia and bipolar disorder. Since they are also being used as outcome measures in pharmacological and cognitive remediation studies aimed at cognitive impairments in schizophrenia, understanding their measurement properties and potential sensitivity to change is important. In this study, the test-retest reliability, practice effects, and reliable change indices of two different performance-based functional capacity measures, the UCSD Performance-based skills assessment (UPSA) and Social skills performance assessment (SSPA) were examined over several different retest intervals in two different samples of people with schizophrenia (n 's=238 and 116) and a healthy comparison sample (n =109). These psychometric properties were compared to those of a neuropsychological assessment battery. Test-retest reliabilities of the long form of the UPSA ranged from $r=.63$ to $r=.80$ over follow-up periods up to 36 months in people with schizophrenia, while brief UPSA reliabilities ranged from $r=.66$ to $r=.81$. Test-retest reliability of the NP performance scores ranged from $r=.77$ to $r=.79$. Test-retest reliabilities of the UPSA were lower in healthy controls, while NP performance was slightly more reliable. SSPA test-retest reliability was lower. Practice effect sizes ranged from .05 to .16 for the UPSA and .07 to .19 for the NP assessment in patients, with HC having more practice effects. Reliable change intervals were consistent across NP and both FC measures, indicating equal potential for detection of change. These performance-based measures of functional capacity appear to have similar potential to be sensitive to change compared to NP performance in people with schizophrenia.

The use of performance-based assessments of functional abilities is becoming more widespread within the field of schizophrenia research (Harvey et al., 2007). These measures directly

© 2010 Elsevier B.V. All rights reserved.

Correspondence to: Philip D. Harvey, PhD, Department of Psychiatry and Behavioral Science, Emory University School of Medicine, Atlanta, GA 30322, 404-727-2707 (Voice), 404-727-3233 (Fax), Philipdharvey1@cs.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

All other authors report no potential competing interests.

examine skills relevant to everyday living abilities, self-care and social outcomes. Performance-based measures of functional ability can often provide a more direct assessment of functional capacity without the influence of environmental and social factors that may influence real-world functioning (Moore et al., 2006). These measures are also quite consistently associated with performance on neuropsychological tests. For example, across 6 different studies, the correlation between scores on the UCSD performance-based skills assessment (UPSA; Patterson et al., 2001a; Mausbach et al., 2007) in both standard and abbreviated versions and Neuropsychological (NP) performance has ranged from $r=.58$ to $r=.65$ (Leifker et al., 2009).

With the advent of the MATRICS initiative (Green et al., 2004; Marder and Fenton, 2004) and increased interest in cognitive enhancement in schizophrenia (Harvey, 2009), these measures have been proposed for use as co-primary measures in cognitive enhancement studies in order to provide evidence regarding the potential functional impact of cognitive changes (Green et al., 2008). Relative to neuropsychological test performance, however, there is markedly less understanding of the psychometric characteristics of these tests. Most test-retest data to date have been presented from very short term studies, such as 4 to 6 weeks (e.g., Green et al., 2008; Harvey et al., 2006), while the MATRICS research design for cognitive enhancement studies is of generally longer duration, such as 6 months (Buchanan et al., 2005). There are currently no published data on the long-term stability, sensitivity to practice effects, and potential sensitivity to change of these performance-based measures.

In this study we present data from two separate longer-term studies of performance-based outcomes in people with schizophrenia. Both of these studies included assessments of 6 months or longer in duration, with follow-up intervals ranging up to 36 months. Two versions of the UPSA were examined (Full and Brief) and both studies also used a performance-based measure of social competence, the Social Skills Performance Assessment (SSPA; Patterson et al., 2001b). In one study, healthy controls were also assessed and a comprehensive neuropsychological assessment was performed. Thus, this study presents long-term psychometric data on performance-based functional capacity assessments of patients with schizophrenia who were observed naturalistically while receiving treatment as usual and in patients with schizophrenia who received an inactive treatment in a clinical trial of a behavioral intervention (Patterson et al., 2006). Neuropsychological test performance and the performance of healthy controls on the same measures is provided to allow for comparison of the psychometric characteristics of these performance-based outcomes measures in schizophrenia against controls using a longer follow-up duration compared to any studies published to date.

Methods

Subjects from two completely separate studies at different research sites were recruited. Both sub-samples were receiving treatment as usual, one sample in a naturalistic follow-up study of cognitive changes with aging in a clinical and control sample (Study 1) and the other sample consisting of patients randomized to treatment as usual in a functional outcomes treatment study (Study 2).

Subjects: Study 1

Schizophrenia Patients—This sample consisted of 238 older schizophrenia outpatients enrolled in a longitudinal study investigating the course of cognitive and functional status. All patients met diagnostic criteria for schizophrenia or schizoaffective disorder. Diagnosis was confirmed by the Comprehensive Assessment of Symptoms and History (CASH; Andreasen et al., 1992). Patients were excluded if they met any of the following criteria: a) a primary DSM-IV Axis I diagnosis other than schizophrenia or schizoaffective disorder, b) Mini-Mental Status Examination (MMSE; Folstein et al., 1975) score below 18, c) Wide Range Achievement Test

(WRAT-3; Jastak, 1994) reading grade-equivalent score of grade 6 or less, or d) any medical illnesses that might interfere with the assessment of cognitive functioning. All patients were receiving antipsychotic treatment over the follow-up period.

If study subjects met the inclusion/exclusion criteria, they were further required to have evidence of continued illness at the time of recruitment, as supported by: a) An inpatient admission for psychosis in the past two years; b) An emergency room visit for psychosis in the past two years; or c) scoring a 4 (moderate) or more on one of three positive symptoms items (delusions, hallucinations, or conceptual disorganization) of the Positive And Negative Syndrome Scale (PANSS; Kay et al., 1987). Outpatient status was defined as living outside of any institutional setting, including a nursing home. All patients were receiving treatment with atypical antipsychotic medications at each assessment. Patients were recruited at Veteran's Affairs Hospital clinics, a New York State Psychiatric Hospital, or Mount Sinai School of Medicine. Upon receiving a description of testing and study procedures, all subjects signed a written informed consent form approved by the institutional review board at each research site. Participants were compensated for time and travel, receiving \$50.00 per assessment.

Healthy Comparison Subjects—Healthy comparison subjects were recruited at a “naturally occurring retirement community (NORC)” in Manhattan. All NORC residents had to be eligible on the basis of lifetime modest income to reside in public housing, which created a population of individuals who were not likely to be extraordinarily high functioning over their lifespan. Other than eligibility for public housing, the only other inclusion criteria for living in the NORC is being retired. All healthy comparison subjects were screened with the CASH and were excluded from participation if they met current or lifetime criteria for major depression or any psychotic condition. They also met the cognitive, reading performance, and substance abuse screening criteria applied to the schizophrenia sample. Healthy controls were also excluded if they were prescribed psychotropic medications for any of the psychiatric conditions that would have led to their exclusion, regardless of whether diagnostic criteria were met with the CASH interview. Healthy control subjects also signed written informed consent approved by the local institutional review board and were also compensated \$50.00 for their time and travel.

Subjects: Sample 2

Schizophrenia Patients—A separate analysis was conducted on data from a previously published study (Patterson et al 2006). Participants were selected from 240 patients recruited from 25 Board and Care facilities in San Diego County. Eligible patients at each facility were recruited for participation. All patients had DSM-IV chart-based diagnoses of schizophrenia or schizoaffective disorder. Patients were excluded if they met any of the following criteria: 1) DSM-IV diagnosis of dementia, 2) were at risk for suicide, 3) could not complete the assessment battery, or 4) were currently participating in a psychosocial intervention or drug research during the intake process. Upon explanation of study procedures and obtaining informed consent, patients were randomized into two groups: one would receive Functional Adaptation Skills Training (FAST) while the other would serve as an attentional control (AC) group.

Patients within the Attention Control group received attention from care staff as usual and also attended 24 weekly group 120-minute group sessions where they could discuss any personal problems or other issues important to them. This control sample was used in the original study to control for clinical contact. Patients in both the AC and FAST condition received \$20 at each assessment session to reimburse them for time and travel. For the purposes of the current study, only data from the participants in the AC cohort was used as these patients did not receive

a rehabilitation-oriented intervention and would thus act as an in-active control group. See Patterson et al., 2006 for additional information on the sample.

Descriptive characteristics on all three samples of research participants are presented in Table 1.

Measures

Performance-based Measures of Functional Capacity—The UCSD Performance-Based Skills Assessment Battery (UPSA; Patterson et al., 2001a) directly assesses functional skills competence among the community dwelling mentally ill. This test measures everyday functioning skills using standardized role-playing situations to evaluate skills in five different functional domains: Comprehension/Planning (e.g., organizing outings to the beach or the zoo), Finance (e.g., counting change, and paying bills), Transportation (e.g., using public transportation), Household Chores (e.g., Planning menus and shopping) and Communication (e.g., using the telephone, rescheduling medication appointments). The test is standardized so that total scores range from 0–100. The participants at the Mt. Sinai site were not assessed with the household chores subtest because the analogue kitchen required was not portable enough to be used at field sites. Thus, Mt. Sinai scores, based on 4 subtests, were converted to a 100-point scale. This modified version was used in previous reports with the UPSA (Bowie et al., 2006; McClure et al., 2007). The UPSA-Brief score was also computed for both patient samples (Mausbach et al., 2007). This UPSA-Brief score contains 2 of the 5 original UPSA domains: Finance and Communication.

The Social Skills Performance Assessment (SSPA; Patterson et al., 2001b) is a role-play measure of social skills that was created for use with schizophrenia patients. After a 1-minute practice period, the patients initiate and maintain a conversation for three minutes in two situations: greeting a new neighbor and calling a landlord to request a repair for a leak that has gone unfixed despite a previous request. These sessions were audio-taped and scored by a trained rater who was unaware of diagnosis (patient or healthy control) and all other data. Patients are scored on a 5-point Likert scale in dimensions of their social skills, including: fluency, clarity, focus, negotiation ability, persistence, and social appropriateness. The blinded raters were trained to the gold standard ratings of the instrument developers (ICC = .86) and high inter-rater reliability was maintained at three months (ICC=.87). Overall scores can be created by summing the scores from each of the domains across situations or creating an overall mean score. For the purposes of this study, an overall mean social functioning score was used and ranged from 1 (low) to 5 (high). The SSPA has previously been shown to be sensitive to the effects of antipsychotic treatment (Harvey et al., 2006), as well as the active FAST intervention described above,

Cognitive Assessment—Research participants at the Mt. Sinai Site (Sample 1) completed a comprehensive assessment of cognitive ability areas previously shown to be correlated with functional skills, including attention, motor skills, verbal learning and memory, verbal fluency, and several aspects of executive functioning. The following tests were included: category fluency (animal naming; Spreen and Strauss, 1998), phonological fluency (FAS; Spreen and Strauss, 1998), Wisconsin Card Sorting Test 64-card computerized version (Heaton et al., 1993) total errors, Trail-making test parts A and B (Reitan and Wolfson, 1993), the Rey Auditory Verbal Learning Test (RAVLT; Spreen and Strauss, 1998) learning trials 1–5, and the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III; Psychological Corporation, 1998) digit span, digit symbol, and letter-number sequencing subtests. We standardized these variables according to previously published norms, creating z-scores, and then averaged all of the cognitive measures to create a single composite score (see Bowie et al., 2008 for a more detailed description of this process).

Follow-up Methods—The Mt. Sinai participants were seen at baseline and two follow-up assessments that were scheduled for 18 and 36 months after the baseline assessments. The UCSD subjects were seen at baseline, 6, 12 and 18 months. The assessment procedures were the same within each cohort at each visit.

Data Analysis

In order to examine changes in performance across time with the procedure most likely to reveal any significant differences, paired samples t-tests were computed to determine the mean difference between baseline scores and each reassessment in the two different samples. For each score we also calculated test-retest reliability by determining the correlation (Pearson's r) of each of the assessments at baseline and follow-up within each sample of participants. Extent of retesting effects was determined by calculation of effect sizes for the changes in performance over time (Cohen's d).

In addition, normative ranges of scores from baseline to the first reassessment for each group were created through computing the reliable change index with a 90% confidence interval. Using the RCI plus practice effect model (See Heaton et al 2001 for a description of this procedure and Harvey et al., 2005 for an application in a different sample of people with schizophrenia), we identified the threshold for identification of reliable changes in performance. Specifically, 90% confidence intervals were developed standard error of the difference (SE_{diff}) for each test. The SE_{diff} describes the spread of the distribution of performance-based change scores that would be expected if no actual change in abilities had occurred. The SE_{diff} was determined for each test using the following formula:

$$SE_{diff} = [(SD_x^2 + SD_y^2)(1 - r_{xy})]^{1/2}$$

Where x is the first assessment and y is the next. The values of SD_x and r_{xx} , were determined from two samples of schizophrenia patients, using baseline to first reassessment results. The practice effect from healthy comparison subjects was used as the "normal" practice effect. Then, a 90% confidence interval for expected retest scores (X_2) was determined by multiplying the SE_{diff} by ± 1.64 , using the formula:

$$90\% \text{ Confidence Interval} = (X_1 + \text{Mean HC Practice Effect}) \pm 1.64 SE_{diff}.$$

Thus, X_1 represents the baseline score for each subject, and the mean practice effect equals the mean of the change scores (retest score minus baseline) for subjects in the HC sample at the same reassessment period. Using this definition, 90% of the retest scores should fall between the lower and upper boundaries (adjusted for practice effects), by chance alone, of this confidence interval. The narrower the re-test confidence interval, the smaller the change in performance on the part on an individual patient would be required to be definitively nonrandom. Large retest confidence intervals, relative to baseline variance in the scores, would suggest that it would be more difficult to detect reliable changes in performance with that particular assessment instrument as an outcome measure and the re-test intervals, the smaller a change score would have to be in order to be definitively non-random. These RCI results are presented for the two patient samples only, because we are interested in the tests' sensitivity to changes on the part of people with schizophrenia.

Results

Demographic characteristics of the three subject samples are presented in table 1. As we were interested in the characteristics of the UPSA across slightly different samples of people with

schizophrenia, we made no attempt to match or select the people with schizophrenia on demographic factors such as age and education. We computed Pearson correlations between age and education with baseline performance on the UPSA and SSPA in the larger Mt. Sinai sample, finding a correlation of $r = -.26$, $p < .001$ for the correlation of age and baseline UPSA scores, $r = -.24$, $p < .001$ for age and SSPA scores, $r = .34$, $p < .001$ for education and UPSA scores, and $r = .12$, $p = .06$ for education and SSPA scores.

Table 2 presents the baseline scores on the measures for the three samples of research participants. We have chosen to present the mean for the entire sample of cases seen at each assessment (which was less than the baseline sample for the Mt Sinai patients), in order to simplify the presentation of the data, as different subsamples are compared at baseline vs. first reassessment and baseline vs. second (and third, if applicable) reassessment. T-tests on the Mt. Sinai sample, comparing the baseline scores on the four performance-based measures with patients who had data at all three assessments to those who had missing data, were non-significant, all $t(237) < 1.12$, all p values > 0.26 .

Table 3 presents the changes over time, including t-tests and effect sizes, as well as test-retest stability for the 4 performance-based variables in the three subject samples. For the long form of the UPSA, the largest change with reassessment was seen in the UCSD patients at the first reassessment, while most changes in the Mt. Sinai sample were very modest and all were non-significant. Test-retest reliability of the long form of the UPSA was over $r = .60$ for all reassessments for the schizophrenia sample, with lower test-retest reliability for the HC sample likely attributable to reduced baseline variance associated with ceiling effects. The UPSA-B had a similar pattern in the schizophrenia patients, in that all test-retest correlations were greater than $r = .65$, with the largest test-retest effect size at $d = .16$. For the HC sample, there was a statistically significant practice effect on the UPSA-B, with an effect size for improvement of $d = .30$.

Test-retest reliability of the SSPA was lower for all participant groups than the UPSA. Cognitive performance manifested test-retest stability that was essentially equivalent to the UPSA coefficients for the patient samples, while the HC sample showed higher test-retest reliability and significant practice effects at both reassessments. Interestingly, there was a significant decline in performance at retest in the Mt. Sinai schizophrenia sample, reflecting an absence of practice effects and possible cognitive decline in this older sample of patients with schizophrenia.

RCI Intervals

For the Mt. Sinai patient sample, the RCI intervals were very similar for UPSA, UPSA-B, and NP performance. Further, the UCSD patients had RCIs that were essentially identical to those seen in the Mt. Sinai patients. The SSPA RCI intervals were essentially the same in the two patient samples, and notably wider than the confidence intervals for the two versions of the UPSA. For changes on the SSPA to be reliable for an individual patient, they would have to be close to 1.5 SD. However, given the size of the SD for the SSPA (0.7 points in each group), changes of less than this RCI interval would be less than a point ($1.4 * 0.7 = 0.98$). This means that a one-point change on the SSPA would be reliably different from baseline scores in the both of the two schizophrenic patient samples. Changes of less than one point are below the resolution level of the scale and the relatively truncated range of possible total scores is likely responsible for the reduced test-retest reliability and relatively wider reliable change index.

Discussion

The present study provides information regarding the temporal stability of functional outcome measures. The value of this information lies in the comparison of NP performance and

functional capacity measures and consideration of “normal” practice effects obtained from healthy controls. Overall the test-retest reliabilities for both versions of the UPSA were quite high, and very similar despite their differences in length, in the schizophrenic patients from both samples. It is important to note that, as the UCSD sample was recruited from residential facilities for a study attempting to reduce disability, the sample represents a more disabled population with poorer performance on the UPSA and UPSA-B at baseline. Furthermore, that sample contained more variability in their levels of symptomatology at baseline as well and their treatment was not monitored as closely as the Mt. Sinai patients. These factors could influence the 18-month test-retest reliabilities with this sample, which were slightly smaller than the Mt. Sinai Sample. Also, inclusion of younger patients might lead to some differences in the stability of these indices.

In contrast, the test-retest reliability of the Healthy Control cohort of Sample 1 was lower on both versions of the UPSA. However, this is most likely due to the lower range of scores seen on these measures as they are expected to have ceiling effects in HCs, given that the UPSA measures disability and is meant for cognitively disabled populations. Likewise, scores of the HCs on the SSPA would not meet criteria for suitable reliability at either reassessment period. Most likely this is also due to the ceiling effect on the SSPA, and the resulting reduced range of possible scores. Note that the SSPA is aimed at the detection of impairments in social competence, which are generally not seen to this degree in healthy samples.

Test-retest reliability of the SSPA scores in the schizophrenic samples were reasonable, although generally less so than the reliabilities seen with the UPSA. Again, the UCSD sample had lower reliability than the Mt. Sinai sample, probably attributable to greater disability in that sample. The SSPA has a reduced range of possible scores, ranging from 1–5, which means that the standard deviations will be higher and the RCIs subsequently will be wider. As such the SSPA might also be expected to be less well correlated with other concurrent variables such as clinical ratings of real-world functional outcomes. This exact finding has been reported in a very recent study (e.g. Leifker et al., 2009). But, as noted above, a change of 1 point, the smallest detectable change for an individual subject, would be definitely non-random.

Previous studies have reported test-retest reliability of the full UPSA at $r=.93$ after a two week interval (c.f. Harvey et al., 2007). However, no studies to date have reported reliability past a 4-week period. The present study shows that the full version of the UPSA generally maintains a reliability of above .70 for up to 36 months with minimal practice effects. With research programs such as MATRICS and TURNS, the field of schizophrenia is increasingly focusing on cognition. Performance-based functional outcome measures provide a unique measure of functionally relevant cognition in that they eliminate the abstract nature of neuropsychological tests and examine a patient’s actual ability to perform the every-day tasks that they would be required to do in the real world. Further, these tests appear to mediate the correlation between NP performance and real-world outcomes in people with schizophrenia (Bowie et al., 2006; 2008) and other severe mental disorders such as bipolar disorder (Mausbach et al., 2010). As long-term treatment trials aimed at cognitive functioning are conducted, the necessity for long term stability and reliability data for performance-based outcome measures grows. This study provides a first step in demonstrating that these performance-based measures possess adequate long-term psychometric properties and as a result have the potential to detect long-term changes in cognitive functioning and related functional abilities. At the same time, these findings indicate that for three different performance-based functionally relevant measures in people with schizophrenia, relatively large changes would be required to be definitively not due to practice effects and random retest variation.

Acknowledgments

This research was supported by NIMH grants number RO1MH63116 and RO1MH 78775 to Dr. Harvey and NIMH Grant number RO1MH078737 to Dr. Patterson.

In the past year, Dr. Harvey has served as a consultant to Solvay Pharma, Dainippon Sumitomo America, Shire Pharma, Merck and Company, and Eli Lilly. He has current grant support from Astra-Zeneca Pharmaceuticals.

Dr. Bowie has received research funding from Johnson and Johnson PRD.

References

- Andreasen NC, Flaum M, Arndt S. The comprehensive assessment of symptoms and history (CASH): An instrument for assessing psychopathology and diagnosis. *Arch Gen Psychiatry* 1992;49:615–623.
- Bowie CR, Leung WW, Reichenberg A, McClure MM, Patterson TL, Heaton RK, et al. Predicting schizophrenia patients' real-world behavior with specific neuropsychological and functional capacity measures. *Biol Psychiatry* 2008;63:505–511. [PubMed: 17662256]
- Bowie CR, Reichenberg A, McClure MM, Harvey PD. Age-Associated Differences in Cognitive Performance in Older Community Dwelling Schizophrenia Patients: Differential Sensitivity of Neuropsychological and Information Processing Tests? *Schizophr Res* 2008;106:50–58. [PubMed: 18053687]
- 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. [PubMed: 16513862]
- Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;12:189–198. [PubMed: 1202204]
- Green MF, Nuechterlein KH, Gold JM, Barch DM, Cohen J, Essock S, et al. 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. [PubMed: 15336511]
- Green MF, Nuechterlein KH, Kern RS, Baade LE, Fenton WS, Gold JM, et al. Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS Psychometric and Standardization Study. *Am J Psychiatry* 2008;16:214–220. [PubMed: 18172018]
- Harvey PD. Pharmacological cognitive enhancement in schizophrenia. *Neuropsychol Rev* 2009;19:324–335. [PubMed: 19507034]
- Harvey PD, Bellack AS, Velligan DI. Performance-Based Measures of Functional Skills: Usefulness in Clinical Treatment Studies. *Schizophr Bull* 2007;33:1138–1148. [PubMed: 17493956]
- Harvey PD, Patterson TL, Potter LS, Zhong K, Brecher M. Improvement in Social Competence with Short-Term Atypical Antipsychotic Treatment: A randomized, double-blind comparison of Quetiapine vs. Risperidone for Social Competence, Social Cognition, and Neuropsychological Functioning. *Am J Psychiatr* 2006;163:1918–1925. [PubMed: 17074943]
- Harvey PD, Palmer BW, Heaton RK, Mohamed S, Kennedy J, Brickman A. Stability of cognitive performance in older patients with schizophrenia: An 8-week test-retest study. *Am J Psychiatry* 2005;162:110–117. [PubMed: 15625208]
- Heaton, RK.; Chellune, CJ.; Talley, JL.; Kay, GG.; Curtiss, G. Wisconsin Card Sorting Test Manual-Revised and expanded. Odessa, FL: Psychological Assessment Resources; 1993.
- Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, Marcotte TD, Grant I. Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Arch Clin Neuropsychology* 2001;16:75–91.
- Kay SR, Fiszbein A, Opler LA. The Positive And Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull* 1987;13:261–276. [PubMed: 3616518]
- Leifker FR, Bowie CR, Harvey PD. Determinants of everyday outcomes in schizophrenia: the influences of cognitive impairment, functional capacity, and symptoms. *Schizophr Res* 2009;115:82–87. [PubMed: 19775869]

- Leifker FR, Patterson TL, Heaton RK, Harvey PD. Validating Measures of Real-World Outcome: The Results of the VALERO Expert Survey and RAND Panel. *Schizophr Bull.* in press.
- Marder SR, Fenton W. 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;72:5–9. [PubMed: 15531402]
- Mausbach BT, Harvey PD, Goldman SR, Jeste DV, Patterson TL. Development of a brief scale of everyday functioning in persons with serious mental illness. *Schizophr Bull* 2007;33:1364–1372. [PubMed: 17341468]
- Mausbach BT, Harvey PD, Pulver AE, Depp CA, Wolyniec PS, Thornquist MH, et al. Relationship of the Brief UCSD Performance-based Skills Assessment (UPSA-B) to Multiple Indicators of Functioning in People with Schizophrenia and Bipolar Disorder *Bipol Disord.* in press.
- McClure MM, Bowie CR, Patterson TL, Heaton RK, Weaver C, Anderson H, et al. Correlations of functional capacity and neuropsychological performance in older patients with schizophrenia: evidence for specificity of relationships? *Schizophr Res* 2007;89:330–338. [PubMed: 16982175]
- Moore DJ, Palmer BW, Patterson TL, Jeste DV. A review of performance-based measures of everyday functioning. *J Psychiatr Res* 2007;41:97–118. [PubMed: 16360706]
- 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* 2001a;27:235–245. [PubMed: 11354591]
- Patterson TL, Mausbach BT, McKibbin C, Goldman S, Bucardo J, Jeste DV. Functional Adaptation Skills Training (FAST): A randomized trial of a psychosocial intervention for middle-aged and older patients with chronic psychotic disorders. *Schizophr Res* 2006;86:291–299. [PubMed: 16814526]
- Patterson TL, Moscona S, McKibbin CL, Davidson K, Jeste DV. Social skills performance assessment among older patients with schizophrenia. *Schizophr Res* 2001b;48:351–360. [PubMed: 11295387]
- Psychological Corporation. WAIS-III and WMS-III Technical Manual. San Antonio, TX: Author; 1998.
- Reitan, RM.; Wolfson, D. The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation. 2nd edition. Tucson, AZ: Neuropsychology Press; 1993.
- Spren, O.; Strauss, E. A compendium of neuropsychological tests and norms. 2ed.. New York: Oxford University Press; 1998.
- Wilkinson, GS. The Wide Range Achievement Test. Wilmington, DE: Wide Range, Inc.; 1993.

Table 1

Demographic characteristics of the Subject Samples

	Mt Sinai – Patients	Mt Sinai – Controls	UCSD Patients
Age (M[SD])	56.99 (9.00)	68.01 (11.81)	49.2 (6.9)
Gender %			
Female	27	53	31
Male	73	47	69
Years of Education (M[SD])	11.86 (4.42)	13.19 (2.19)	11.9 (2.9)
PANSS Score [M(SD)]	50.14 (10.83)	--	62.5 (17.4)

Table 2

Scores on the Performance Based Measures over the follow-up period

	Baseline		18 Months		36 Months	
	N	M (SD)	N	M (SD)	N	M (SD)
Mt Sinai Sample						
UPSA (Full)						
Patients	238	72.2 (18.3)	136	74.6 (16.4)	69	73.2 (18.3)
Controls	109	84.6 (8.5)	76	87.0 (7.7)	53	86.4 (7.0)
UPSA Brief						
Patients	238	68.65 (20.25)	136	68.64 (20.30)	69	68.10 (20.79)
Controls	109	84.5 (9.0)	76	87.2 (7.6)	53	87.4 (7.7)
SSPA						
Patients	235	3.9 (0.7)	140	3.9 (0.7)	72	3.8 (0.8)
Controls	108	4.5 (0.5)	76	4.6 (0.4)	52	4.7 (0.4)
Neuropsych						
Patients	239	-1.5 (1.0)	136	-1.5 (1.0)	69	-1.6 (1.0)
Controls	94	-0.3 (0.8)	76	-0.02 (0.8)	53	0.1 (0.7)
UCSD Sample						
	Baseline		6 Months		12 Months	
	N	M (SD)	N	M (SD)	N	M (SD)
UPSA (full)	116	64.4 (19.3)	101	67.4 (16.2)	87	66.0 (17.0)
						85 66.3 (18.7)
UPSA-B	116	56.5 (23.0)	101	58.8 (19.5)	87	56.8 (20.5)
						85 55.3 (22.3)
SSPA	116	3.3 (0.73)	102	3.4 (0.68)	89	3.6 (0.60)
						84 3.5 (0.63)

Table 3

Paired t-tests, test-retest reliability, and effect sizes for Change

	Baseline vs. 18 months			Baseline vs. 36 Months					
	Paired t-test			Paired t-test					
	N	T (ES)	r	N	T (ES)	r	N	T (ES)	R
Mt. Sinai Sample									
UPSA (Full)									
Patients	133	0.82 (.05)	.75**	68	1.74 (.11)	.73**			
Controls	109	0.60 (.07)	.52**	53	0.64 (.07)	.53**			
UPSA Brief									
Patients	133	0.01 (.01)	.75**	68	-1.71 (.15)	.73**			
Controls	109	4.64** (.30)	.58**	53	0.35 (.02)	.59**			
SSPA									
Patients	137	0.44 (.03)	.67**	71	-2.41* (0.21)	.78**			
Controls	76	0.61 (.09)	.53**	52	2.24* (0.39)	.41*			
NP Composite									
Patients	132	-1.08 (.07)	.77**	69	-2.28* (0.19)	.79**			
Controls	71	4.44** (.26)	.86**	51	3.90** (.36)	.76**			
UCSD Sample									
UPSA (full)	101	3.20** (.16)	.77**	87	2.60* (.08)	.80**	85	2.43 (.10)	.63**
UPSA-B	101	2.28* (.16)	.79**	87	1.50 (.10)	.81**	85	0.68 (.05)	.66**
SSPA	102	1.13 (.14)	.55**	89	4.22** (.41)	0.55**	84	3.36** (.27)	.49**

* p<.05

** p<.01

Table 4

Levels of change that would be required to be definitely nonrandom in Raw Score and Standard Deviation Units

	Number of Subjects	Level of change to Exceed 90% CI	
Mt. Sinai Sample	N	Raw Score Change required	Level of Change in SD Units
UPSA	133	19.67	1.07
UPSA-B	133	21.32	1.06
SSPA	137	0.88	1.40
Neuropsych	132	1.18	1.18
UCSD Sample	N	Raw Score Change required	Level of Change in SD Units
UPSA (full)	101	21.92	1.13
UPSA-B	101	25.06	1.09
SSPA	102	1.08	1.47