Published in final edited form as:

Am J Psychiatry. 2011 November ; 168(11): 1195–1201. doi:10.1176/appi.ajp.2011.10121723.

Validating the Measurement of Real-World Functional Outcomes: Phase I Results of the VALERO Study

Philip D. Harvey, PhD^{1,2}, Tenko Raykov, PhD³, Elizabeth Twamley, PhD⁴, Lea Vella, MA, MPH⁵, Robert K. Heaton, PhD⁴, and Thomas L. Patterson, PhD⁴

¹University of Miami Miller School of Medicine

²Research Service, Miami VA Medical Center

³Michigan State University

⁴UCSD Department of Psychiatry

⁵SDSU/UCSD Joint Doctoral Program in Clinical Psychology

Abstract

Objective—Cognitive deficits are associated with disability in people with schizophrenia so treatment of cognitive impairment has been proposed as an intervention to reduce disability. However, studies relying on patient self-report have found very minimal relationships between ratings of real-world functioning and cognitive performance, raising questions about the measurement of real-world functioning as a treatment outcome. The Validation of Everyday Real-world Outcomes (VALERO) study was conducted to evaluate functional rating scales and to identify the rating scale or scales most robustly related to performance-based measures of cognition and everyday living skills.

Method—198 adults with schizophrenia were tested with the neurocognitive measures from the MATRICS Consensus cognitive Battery and performed the UCSD performance-based skills assessment-Brief and advanced finances subtest from the Everyday Functioning Battery. They and a friend, relative, clinician, or case manager also reported their everyday functioning on 6 ratings scales: Social Behavior Schedule, Social Adjustment Scale, Heinrichs Carpenter Quality of Life Scale, Specific Levels of Functioning, Independent Living skills Survey, and Life Skills Profile. Best judgment ratings were generated by an interviewer who administered the rating scales to patients and informants.

Results—Statistical analyses developed an ability latent trait that reflected scores on the three performance-based (i.e., ability) measures and canonical correlation analysis related interviewer ratings to the latent trait. The overall fit of the model with all six rating scales was good: $\chi^2 = 78.100$, $df = 56$, $p = .027$, and $RMSEA = .078$. Individual rating scales that did not improve the fit of the model were systematically deleted and a final model with two rating scales fit the data: $\chi^2 = 32.059$, $df = 24$, $p = .126$, $RMSEA = .072$. A regression analysis found that the Specific Levels of Functioning was a superior predictor of the three-performance based ability measures.

Conclusions—We found that systematic assessments of real world functioning are related to performance on neurocognitive and functional capacity measures. Of the six rating scales

Correspondence to: Philip D. Harvey, PhD, Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, 1120 NW 14th Street, Suite 1450, Miami, FL 33136, 305-243-1619 (Fax), philipdharvey1@cs.com.

Dr. Harvey has received consulting fees from Abbott Labs, Bristol Myers Squibb, Cypress Bioscience, En Vivo, Genentech, Merck and Company, Shire Pharma, Sunovion Pharma, and Teva Pharma, during the past year. None of the other authors have any commercial interests to report. This study is unrelated to Dr. Harvey's Consulting work.

evaluated, the Specific Levels of Functioning (SLOF) was best in this study. Use of a single rating scale provides a very efficient assessment of real-world functioning that accounts for considerable variance in performance-based scores.

Disability in multiple everyday functional domains is common in people with schizophrenia [1], encompassing social [2], vocational [3], and residential [4] domains. These impairments are present in the majority of patients and are refractory to current pharmacological interventions [5] and only partially responsive to rehabilitation [6]. Some of the patient characteristics associated with these impairments include cognitive deficits [7], negative symptoms [8], and impairments in performance of social and daily living skills (i.e., “functional capacity” [9]). There are also environmental contributions to real-world functional deficits, which include disability compensation (associated with reduced vocational achievement [10]), and lack of social opportunities and community resources). Disability, as a consequence, is a complex and multiply determined phenomenon which may have similarities across different severe mental illnesses [11].

There have been considerable efforts to treat the putative causes of these functional deficits, including pharmacological interventions aimed at cognitive enhancement and reduction of negative symptoms [5], cognitive remediation interventions [12], supported employment [6] and various psychosocial interventions, including efforts to teach social and everyday living skills [13]. These efforts have met with variable success, with combined interventions including cognitive remediation and psychosocial interventions yielding the most success [14]. However, several studies have found that impairments on a variety of ability measures, including both neuropsychological tests and measures of functional capacity, were weakly related to ratings of real-world functioning [15–17]. These relationships raise the legitimate question as to whether improving cognitive or functional abilities have the potential to exert a meaningful influence on real-world outcomes.

As we have suggested before [18], many studies finding low correlations between ability or capacity measures and ratings of real world outcomes used self-reports of real-world functioning. As shown in several systematic comparisons, patient self reports of both ability (e.g., cognitive performance) and real-world functioning are remarkably unrelated to performance on performance-based ability measures and to informant ratings of ability and real-world functioning. As a result of these findings, the Validation of Everyday Real-World Outcomes (VALERO) initiative was undertaken. The current study in the VALERO initiative is aimed at a direct comparison of the relationship of a set of instruments for rating of real-world functional outcome, by informants or patients themselves, and the ability measures, including both cognitive and functional capacity assessments of people with schizophrenia.

In the present study, six rating scales assessing real-world functional outcomes, previously recommended as the best available by a RAND expert panel, were completed (a) by patients with schizophrenia, (b) an informant (friend, relative, or high-contact clinician), and (c) by a research examiner who conducted the interviews with the patient and informant. The interviewer’s ratings reflected her best estimate of the level of functioning of the patient. Examiners were instructed to base all judgments on what they thought was correct, including discounting any information that they believed was inaccurate. These research examiner ratings were then related to the patient’s performance on the MATRICS Consensus Cognitive Battery (MCCB [19]) and two functional capacity measures, the UCSD Performance-based Skills Assessment-Brief Version (UPSA-B [20]) and the Advanced Finances Subscale of the Everyday Functioning Battery (EFB [21]). The UPSA-B is widely used in studies of functional capacity in schizophrenia, but the EFB is aimed at higher functioning individuals and was included in order to avoid the possibility of ceiling

effects in the assessment of higher functioning patients. The goal of this study was to identify the rating scale or scales that measure real-world functioning most strongly correlated with patients' performance on measures of their ability: cognitive performance and functional capacity. Using statistical techniques described in the on-line supplement, we examined the 6 different scales and found the one most strongly correlated with the performance-based measures.

Method

Participants

The study participants were patients with schizophrenia who were receiving treatment at one of three different outpatient service delivery systems, two in Atlanta and one in San Diego. In addition, informants were interviewed concerning the everyday functioning of each of the patients, with these informants either being a high-contact clinician (case manager, psychiatrist, therapist, or residential facility manager; 20% of cases) or a friend or relative (80% of cases). All of these research participants provided signed, informed consent, and this research study was approved by local IRBs in Atlanta and San Diego. In Atlanta, patients were either recruited at a psychiatric rehabilitation program (Skyland Trail) or from the general outpatient population of the Atlanta VA Medical Center. The Skyland Trail patients were receiving treatment because of functional disability, including impairments in both residential and vocational functioning, and these patients were recruited through their case managers. The Atlanta VA Medical Center patients were not selected for the presence of disability and were recruited through advertisements, from another research project, and by word of mouth. The San Diego patients were recruited from the UCSD Outpatient Psychiatric Services clinic, a large public mental health clinic, and other local community clinics and by word of mouth.

All patients were administered a structured diagnostic interview, either the Structured Clinical Interview for the DSM (SCID; [22] administered at the Atlanta sites) or the Mini International Neuropsychiatric Interview, 6th Edition (MINI [23] administered at the San Diego site) by a trained interviewer. All diagnoses were subjected to a consensus procedure at each local site. Patients were excluded for a history of traumatic brain injury with unconsciousness >10 minutes, brain disease such as seizure disorder or neurodegenerative condition, or the presence of another DSM-IV diagnosis that would exclude the diagnosis of schizophrenia. None of the patients were experiencing their first psychiatric admission. Comorbid substance use disorders were not an exclusion criterion, in order to capture a broad array of patients, but patients who appeared intoxicated were rescheduled. Inpatients were not recruited, but patients resided in a wide array of unsupported, supported, or supervised residential facilities. Informants were not screened for psychopathology or substance abuse. Descriptive information on patients and informants is presented in Table 1.

Procedure

All patients were examined with a performance-based assessment of neurocognitive abilities and functional capacity. They also provided self-reports of social, residential, and vocational functioning on six different functional outcomes scales either administered as interviews by a trained rater or completed as a questionnaire. Informants independently completed the same six outcomes scales reporting on the functioning of the patients. The examiner who conducted the interviews with the patient and informant then generated ratings for all six rating scales, based on her impression of the "true" status of the patient. Rating scales were presented in a fixed, counterbalanced order across patients. Clinical ratings of symptoms were also collected for descriptive purposes and are presented in Table 1, along with demographic information.

Performance-based assessment

Neurocognition—We examined cognitive performance with a modified version of the MATRICS consensus cognitive battery (MCCB [19]). For this study, we did not include the social cognition measure, from the MCCB, the Mayer–Salovey–Caruso Emotional Intelligence Test Managing Emotions, because there are several reasons to think that social cognition measures may have a different relationship with everyday outcomes compared to neurocognitive measures. This minor modification of the MCCB would make the results similar to previous work, such as our own [3,11] that did not include social cognition measures. We calculated a composite score, an average of 9 age-corrected T-scores based on the MCCB normative program, as our critical dependent variable.

Functional Capacity—We administered two different performance-based functional capacity measures. Participants' functional abilities were assessed using the Brief version of the UCSD Performance-based Skills Assessment (UPSA-B [20]). The UPSA-B is a measure of functional capacity in which patients are asked to perform everyday tasks related to communication and finances. During the *Communication* subtest, participants role-play exercises using an unplugged telephone (e.g., emergency call; dialing a number from memory; calling to reschedule a doctor's appointment). For the *Finance* subtest, participants count change, read a utility bill, and write and record a check for the bill. The UPSA-B requires approximately 10–15 minutes, and raw scores are converted into a total score ranging from 0–100, with higher scores indicating better functional capacity. We also administered the Advanced Finances subscale of the Everyday Functioning Battery (EFB [21]), designed to examine financial management in higher functioning individuals. The Advanced Finance test requires individuals to prepare bank deposits and checks to pay bills, maintain a checkbook balance, and organize payments such that a pre-specified amount of money is left available at the end of the task. This instrument was selected because it measures abilities considered important for independent living and at the time the study was planned, we were concerned that younger individuals with schizophrenia might evidence ceiling effects on the UPSA-B. Total scores on the Advances Finances subtest range from 0–13.

Real-World Functional Outcomes—As we previously reported [18], the initial phase of the VALERO study included a RAND panel that selected 6 functional outcomes scales from a much larger group of candidate scales as most suitable for current use at the time of the panel (see [18] for detailed descriptions of these instruments). These six scales are the Heinrichs-Carpenter Quality of Life Scale (QLS [24]), Specific Levels of Functioning (SLOF [25]), Social-Behavior Schedule (SBS [26]), Social Functioning Scale: (SFS [27]), Life Skills Profile (LSP [28]), and the Independent Living Skills Survey (ILSS [29]).

There are several important features of these functional scales. Two (SBS, SFS) were pure social functioning scales, while two others examined only community functioning (LSP; ILSS). The others (QLS; SLOF) were “hybrid” scales examining social, residential, and vocational outcomes. Of the six scales, 2 were administered as self-report questionnaires (ILSS and SLOF) and the others were administered as interviews using the standard instructions for the scale. Although all of these rating scales have multiple individual subscales, for the purposes of this first report, we examined only total scores. If these scales were used as outcomes measures in a clinical trial, a single, predefined primary outcome would be selected and we were interested in making this information available as straightforwardly as possible.

Some of these instruments were modified by deletion of some subscales following the suggestions of the RAND panel. For instance, the social acceptability and personal care

subscales of the SLOF were not used for the total SLOF scores and for the QLS, the intrapsychic foundations subscale was not included in the analyses because it measures deficit (i.e., negative) symptoms. While negative symptoms are known to affect functional outcomes [11], we were interested in examining the association between performance-based measures and functioning. Similar to our decision to exclude social cognition from the neurocognitive predictor set, we wanted to exclude negative symptoms as an outcome measure.

Data Analysis

The primary goal of the data analysis was to find the real world functional outcome scale, or scales, that was most strongly related to the three performance-based indices of functional ability on the part of the patients. This analysis was conducted using the method of robust maximum likelihood parameter estimation, which included all available data and did not assume multivariate normality for observed measures [30]. With this analytic approach, the model is fitted to all available data from all subjects. That is, *no* subject is dropped from any of the analyses due to not providing data on more than one of the variables in any model.

To accomplish these aims, a structural equation model was developed and fit to the available data from the three sites with the latent variable modeling software Mplus [30]. A single latent trait reflecting the shared variance of the three performance-based “ability” variables was developed using Hierarchical linear modeling. This single trait was then statistically related to examiner-generated total scores on all six of the real-world functional outcomes scales. This overall model with a single ability latent trait and 6 scales as predictors was tested for its goodness of fit with standard indices, including χ^2 (chi-square value), degrees of freedom, associated *p*-value, and root mean square error of approximation (RMSEA). As discussed in the structural equation modeling and confirmatory factor analysis literature [31], smaller scores on both χ^2 and RMSEA are indicators of desirable fit (useful approximation to the analyzed data). Further, shared variance statistics are calculated with an R^2 statistic for the shared multiple correlation between real-world functioning measures and the performance-based ability latent trait.

Following the fitting of this model, the real-world outcomes variables were considered for deletion from the model in a sequential order based on the lowest correlations for their loadings on the ability latent trait. After deletion of the real-world functional scale with smallest correlation, the overall model fit was recalculated and if there was room for improvement in the fit, the next scale with a low correlation was deleted. As the dimension of real-world functioning needed to be defined by at least two rating scales, a total of 4 of the 6 scales could be considered for deletion and the final two could be compared with regression analysis for their relative predictability of the ability latent trait.

Results

Demographic, clinical, and treatment characteristics of the patients are presented in Table 1. As can be seen in the table, there were demographic differences between the sites. These included age, education, and racial differences, differences in ethnicity, and in the type of informant. All of the informants at the Atlanta VA were friends or relatives. Further, there were differences in residential status, with the majority of the San Diego and Atlanta VA patients were living in the community; the Skyland Trail patients were more residentially disabled. There were no differences in employment across the samples and the PANNS scores were essentially identical across sites. The scores on the three performance-based measures were very similar across the sites as well, although the Atlanta VA and UCSD patients had somewhat higher UPSA scores. Thus, the demographic differences between the

samples and sites did not correspond to major differences in the performance-based measures.

Table 2 presents the overall fit of the complete baseline model with the ability latent trait based on three performance-based indicators and examiner-generated total scores for the 6 real-world functioning scales. As can be seen in the top half of the table, all three performance-based variables were significantly related to the ability latent trait. As seen in the bottom of the table, only one of the 6 rating scales was significantly related to the ability latent trait: the Specific Levels of Functioning in the baseline analyses. Nonetheless, the overall fit of the model still would be considered to be acceptable, in that the RMSEA was small and the p -value close to nonsignificant. However, this model clearly had room for improvement, in that most of the rating scales were not independently related to the performance-based measures.

As a result, we dropped real-world functioning scales in sequence, as presented in Figure 1. The sequence was determined by a rank ordering of the nonsignificant p -values for the factor loadings between the real-world rating measure and the ability latent trait. The order of deletion was Social Behavior Schedule, Independent Living Skills Survey, Quality of Life Scale, and Social Functioning Scale based on these criteria. In each of these models based on sequential deletion, there was still room for improvement in the fit of the model, although some began to show evidence of fitting the data acceptably (e.g., the model with Social Behavior Schedule and Independent Living Skills Survey deleted). When four of the 6 scales had been deleted, the model judged as best fitting was revealed, which suggested that the Life Skills Profile and Specific Levels of Functioning were the two best scales in combination for the prediction of the ability latent trait.

The Life Skills Profile factor coefficient was not statistically significant in the last and best fitting model, suggesting that it did not add any information to that provided by the Specific Levels of Functioning. For this reason, a simultaneous regression analysis was performed, entering the two functional outcomes examiner-rated total scores as predictors of the ability latent trait. The analysis found that the Specific Levels of Functioning total score was significantly related to the ability trait, $t(192)=3.09$, $p=.002$, while the Life Skills Profile total score was not, $t(192)=1.41$, $p=.15$. In a forced entry regression analysis, when the Specific Levels of Functioning was entered first, the results were significant, $t(193)=4.52$, $p<.001$, while the LSP total score did not enter, $t(192)=.88$, $p=.38$. When the LSP total score was forced into the regression analysis first, it still did not enter the equation, $t(193)=1.75$, $p=.081$, while the Specific Levels of Functioning still contributed variance to the ability latent trait above and beyond the nonsignificant contribution of the Life Skills Profile, $t(192)=4.21$, $p<.001$. When total variance accounted for between the Ability Latent trait and Specific Levels of Functioning total scores was examined, it was revealed that the Specific Levels of Functioning ratings accounted for 24% of the variance in the ability latent trait. Thus, reducing the number of real-world functional rating scales from 6 to 1 leads to a reduction in reliable variable of 17%, but allowed for the collection of functional information with a single questionnaire.

Discussion

The results of this study indicate that real-world functional outcomes rated with an array of preselected rating scales, utilizing information from the patient and an informant condensed into a judgment rating by an examiner, are related on a global basis to performance-based assessments of ability in people with schizophrenia. This is an important finding, because several different studies reviewed above have found correlations between cognitive and functional abilities and real-world outcomes that were modest to negligible, raising

questions about the well-accepted relationship between cognitive impairment and disability in people with schizophrenia. These data suggest that this relationship is actually as strong as suggested by Green et al. [32], but also indicates that both the rating scale and the methods for rating real-world functioning may influence the strength of this association.

Further, the results indicate that many rating scales addressing real-world functioning, even with multiple sources of information and a systematic approach to ratings in optimal research conditions, are not strongly related to indices of functional abilities that are the state of the art outcomes in treatment studies. Finally, it is possible to tentatively endorse a functional outcomes scale (i.e., the Specific Levels of Functioning) that measures social, everyday living, and vocational outcomes and is related to performance on measures of everyday functional ability at a level that provides substantial information about everyday functioning.

In studies that use these performance-based measures as outcomes, it appears as if the Specific Levels of Functioning Scale could be a suitable baseline measure to index ability-relevant real-world functioning or an outcome measure in a long-term study, as well as for use in clinical assessment. Most of the real-world functional outcomes scales seem to largely be redundant with each other when utilized simultaneously and the results suggest that none of them is related to functional abilities above and beyond the relationship of the SLOF. However, this finding does not prove that other rating scales would not be suitable with similar rating methods. Later analyses will examine the utility the subscales of these rating scales for augmenting the SLOF for assessment of ability-relevant real-world functioning and which, if any, of the informant ratings could be substituted for the comprehensive interviewer judgments.

There are a number of limitations to be considered when evaluating these results. As noted immediately above, it is possible that elements of these real-world functioning scales, on either a subscale or item basis, could also relate more efficiently to ability measures than the total scores. Further, the current analyses did not address whether self or informant reports of functioning could possibly be superior to interviewer judgments and whether either of these sources of information could be safely used on their own with the current group of scales. This question will be addressed in detail in subsequent analyses. There are other real-world functional status scales that were not examined. Most of these scales were excluded at either the scale screening or RAND panel level in the early phases of this study because there was not enough data available to evaluate their suitability. Some scales have been developed since that time, such as the Schizophrenia Objective Functioning Instrument [33], or have had several new studies published, such as the Personal and Social Performance Scale [34]. Finally, since we did not use the MATRICS social cognition measure, we could not compute the composite score with the MCCB normative sample scoring program. This program compares composite scores to those obtained from healthy individuals and leads to composite overall scores in impaired populations that are lower than the average of the individual item scores. For example, in a recently published clinical trial [35], the screening composite t-score on the MCCB was 24.7, but the average of the 9 tests we administered was 34.2. This score is similar to the average score on the same tests in our sample of 38.6, indicating that this is not a relatively unimpaired sample.

As is known from previous studies by our team and others, there are other factors that add to the prediction of real-world outcomes, with some of these factors occasionally found to have a greater impact on real-world functioning than ability (e.g., disability compensation, depression, or negative symptoms). However, even if other factors also predict outcome, treatments aimed at disability reduction try to improve ability, not to change environmental, cultural, or emotional variables influencing disability. Later phases of the VALERO study

will examine the usefulness of different informants for generation of real-world functional ratings on the SLOF and later analyses of the current data set will address issues of site and informant differences in validity of real-world functional ratings. The analyses in this paper are aimed at identification of the most broadly useful real-world outcomes rating scale, when administered to the patient and informant, and followed by a best estimate rating. Such a rating procedure has been successfully implemented in previous treatment studies [36], suggesting that it is a feasible procedure.

Acknowledgments

This research was supported by Grants MH078775 to Dr. Harvey and MH078737 to Dr. Patterson from the National Institute of Mental Health.

References

1. Murray CJL, Lopez AD. Global mortality, disability, and the contributions of risk factors: Global burden of disease study. *Lancet*. 1997; 349:1436–1442. [PubMed: 9164317]
2. Wiersma D, Wanderling J, Dragomirecka E. Social disability in schizophrenia: its development and prediction over 15 years in incidence cohorts in six European centres. *Psychol Med*. 2000; 30:1155–1167. [PubMed: 12027051]
3. Bowie CR, Leung WW, Reichenberg A, McClure MM, Patterson L, Heaton RK, Harvey PD. Predicting schizophrenia patients' real world behavior with specific neuropsychological and functional capacity measures. *Biol Psychiatry*. 2008; 63:505–511. [PubMed: 17662256]
4. Twamley EW, Doshi RR, Nayak GV, Palmer B. Generalized cognitive impairments, ability to perform everyday tasks, and level of independence in community living situations of older patients with psychosis. *Am J Psychiatry*. 2002; 159:2013–2020. [PubMed: 12450950]
5. Harvey PD. Pharmacological cognitive enhancement in schizophrenia. *Neuropsychol Rev*. 2009; 19:324–35. [PubMed: 19507034]
6. Cook JA, Leff HS, Blyler CR, Gold PB, Goldberg RW, Mueser KT, Toprac MG, McFarlane WR, Shafer MS, Blankertz LE, Dudek K, Razzano LA, Grey DD, Burke-Miller J. Results of a multisite randomized trial of supported employment interventions for individuals with severe mental illness. *Arch Gen Psychiatry*. 2005; 62:505–12. [PubMed: 15867103]
7. Green MF, Kern RS, Braff DL, Mintz J. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the “right stuff”? *Schizophr Bull*. 2000; 26:119–136. [PubMed: 10755673]
8. Ventura J, Helleman GS, Thames AD, Koellner V, Nuechterlein KH. Symptoms as mediators of the relationship between neurocognition and functional outcome in schizophrenia: a meta-analysis. *Schizophr Res*. 2009; 113:189–199. [PubMed: 19628375]
9. Bowie CR, Leung WW, Reichenberg A, McClure MM, Patterson L, Heaton RK, Harvey PD. Predicting schizophrenia patients' real world behavior with specific neuropsychological and functional capacity measures. *Biol Psychiatry*. 2008; 63:505–511. [PubMed: 17662256]
10. Rosenheck R, Leslie D, Keefe R, McEvoy J, Swartz M, Perkins D, Stroup S, Hsiao JK, Lieberman J. CATIE Study Investigators Group. Barriers to employment for people with schizophrenia. *Am J Psychiatry*. 2006; 163:411–417. [PubMed: 16513861]
11. Bowie CR, Depp C, McGrath JA, Wolyniec P, Mautsach BT, Thornquist MH, Luke J, Patterson TL, Harvey PD, Pulver AE. Prediction of real-world functional disability in chronic mental disorders: a comparison of schizophrenia and bipolar disorder. *Am J Psychiatry*. 2010 Sep. 167:1116–24. [PubMed: 20478878]
12. McGurk SR, Twamley EW, Sitzer DI, McHugo GJ, Mueser KT. Meta-analysis of cognitive remediation in schizophrenia. *Am J Psychiatry*. 2007; 164:1791–1802. [PubMed: 18056233]
13. Patterson TL, Mautsach BT, McKibbin C, Jeste DV. Functional adaptation skills training (FAST): a randomized trial of psychosocial intervention for middle-aged and older patients with chronic psychotic disorders. *Schizophr Res*. 2006; 86:291–299. [PubMed: 16814526]

14. McGurk SR, Mueser KT, Feldman K, Wolfe R, Pascaris A. Cognitive Training for Supported Employment: 2–3 Year Outcomes of a Randomized Controlled Trial. *Am J Psychiat*. 2007; 164:437–441. [PubMed: 17329468]
15. Green MF, Nuechterlein KH, Kern RS, Baade LE, Fenton WS, Gold JM, Keefe RS, Mesholam-Gately R, Seidman LJ, Stover E, Marder SR. Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS Psychometric and Standardization Study. *Am J Psychiatry*. 2008; 165:221–228. [PubMed: 18172017]
16. McKibbin C, Patterson TL, Jeste DV. Assessing disability in older patients with schizophrenia: results from the WHODAS-II. *J Nerv Ment Dis*. 2004; 192:405–413. [PubMed: 15167403]
17. Bowie CR, Twamley EW, Anderson H, Halpern B, Patterson T, Harvey PD. Self-assessment of functional status in schizophrenia. *J Psychiatr Res*. 2007; 41:1012–1018. [PubMed: 17014866]
18. 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*. (Epub ahead of print, Jun 12, 2009).
19. Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch D, Cohen J, et al. The MATRICS Consensus Cognitive Battery: Part 1. Test selection, reliability, and validity. *Am J Psychiat*. 2008; 165:203–213. [PubMed: 18172019]
20. 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; 33:1364–72. [PubMed: 17341468]
21. Heaton RK, Marcotte TD, Mindt MR, Sadek J, Moore DJ, Bentley H, McCutchan JA, Reicks C, Grant I. HNRC Group. The impact of HIV-associated neuropsychological impairment on everyday functioning. *J Int Neuropsychological Soc*. 2004; 10:317–331.
22. First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW.; Benjamin, LS. User's guide for the Structured Clinical Interview for DSM-IV Axis I (SCID-I). Washington, DC: American Psychiatric Press; 1995.
23. Sheehan DV, Lecrubier Y, Harnett-Sheehan K, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar G. The Mini International Neuropsychiatric Interview (M.I.N.I.): The Development and Validation of a Structured Diagnostic Psychiatric Interview. *J Clin Psychiatry*. 1998; 59(suppl 20): 22–33. [PubMed: 9881538]
24. Heinrichs DW, Hanlon TE, Carpenter WTJ. The Quality of Life Scale: An instrument for rating the schizophrenia deficit syndrome. *Schizophr Bull*. 1984; 10:388–396. [PubMed: 6474101]
25. Schneider LC, Struening EL. SLOF: a behavioral rating scale for assessing the mentally ill. *Soc Work Res Abstr*. 1983; 19:9–21. [PubMed: 10264257]
26. Wykes T, Stuart E. The measurement of social behavior in psychiatric patients: and assessment of the reliability and validity of the SBS schedule. *Br J Psychiatry*. 1986; 148:1–11. [PubMed: 3082403]
27. Birchwood M, Smith J, Cochrane 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 programmes with schizophrenic patients. *Br J Psychiatry*. 1990; 157:853–859. [PubMed: 2289094]
28. Rosen A, Hadzi-Pavlovic D, Parker G. The Life Skills Profile: A measure assessing function and disability in schizophrenia. *Schizophr Bull*. 1989; 15:325–337. [PubMed: 2749191]
29. Wallace CJ, Liberman RP, Tauber R, Wallace J. The Independent Living Skills Survey: A comprehensive measure of the community functioning of severely and persistently mentally ill individuals. *Schizophr Bull*. 2000; 26:631–658. [PubMed: 10993403]
30. Muthen, LK.; Muthen, B. Mplus user's guide. Los Angeles, CA: Muthen & Muthen; 2010.
31. Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing Structural Equation Models*. Beverly Hills, CA: Sage; 1993. p. 136-162.
32. Green MF, Kern RS, Braff DL, Mintz J. Neurocognitive deficits and functional outcome in schizophrenia: Are we measuring the “right stuff”? *Schizophr Bull*. 2000; 26:119–136. [PubMed: 10755673]
33. Kleinman L, Lieberman J, Dube S, Mohs R, Zhao Y, Kinon B, Carpenter W, Harvey PD, Green MF, Keefe RS, Frank L, Bowman L, Revicki DA. Development and psychometric performance of

- the schizophrenia objective functioning instrument: an interviewer administered measure of function. *Schizophr Res.* 2009; 107:275–85. [PubMed: 19013769]
34. Patrick DL, Burns T, Morosini P, Rothman M, Gagnon DD, Wild D, Adriaenssen I. Reliability, validity and ability to detect change of the clinician-rated Personal and Social Performance scale in patients with acute symptoms of schizophrenia. *Curr Med Res Opin.* 2009; 25:325–335. [PubMed: 19192977]
35. Keefe RSE, Fox KH, Harvey PD, Cucchiaro J, Siu C, Loebel A. Characteristics of the MATRICS Consensus Cognitive Battery in a 29-site antipsychotic schizophrenia clinical trial. *Schizophr Res.* 2011; 125:161–168. [PubMed: 21075600]
36. Harvey PD, Ogasa M, Cucchiaro J, Loebel A, Keefe RSE. Performance and Interview-based Assessments of Cognitive Change in a Randomized, Double-blind Comparison of Lurasidone vs. Ziprasidone. *Schizophr Res.* in press.

Original model (baseline model, Table 2)

$\chi^2 = 78.100$, $df = 56$, $p = .027$, RMSEA = 0.078, 90% CI = 0.027–0.117

Social Behavior Schedule deleted

$\chi^2 = 65.764$, $df = 48$, $p = 0.046$, RMSEA = 0.075, 90% CI = 0.012–0.118

Social Behavior Schedule and Independent Living Skills Survey deleted

$\chi^2 = 53.944$, $df = 40$, $p = 0.069$, RMSEA = 0.073, 90% CI = 0.000–0.120

Social Behavior Schedule and Independent Living Skills Survey and Quality of Life Scale deleted

$\chi^2 = 49.275$, $df = 32$, $p = 0.026$, RMSEA = 0.091, 90% CI = 0.032–0.139

Social Behavior Schedule, Independent Living Skills Survey, Quality of Life Scale, and Social Functioning Scale deleted^b

$\chi^2 = 32.059$, $df = 24$, $p = 0.126$, RMSEA = 0.072, 90% CI = 0.000–0.131

^a RMSEA = root mean square error of approximation

^b Best fitting model: smallest chi-square value, largest p-value, and smallest RMSEA; the remaining scales were the Life Skills Profile and the Specific Levels of Functioning Assessment.

Figure 1. Model Fits During Sequential Deletion of Real-world Functioning Scales^a

^a RMSEA = root mean square error of approximation

^b Best fitting model: smallest chi-square value, largest p-value, and smallest RMSEA; the remaining scales were the Life Skills Profile and the Specific Levels of Functioning Assessment.

Table 1
Demographic and clinical characteristics and performance-based scores for participants at the three study sites

Variable	Site								Analysis	
	Skyland Trail (n=55)		Atlanta VA Medical Center (n=40)		UCSD (n=100)					
	n	%	n	%	n	%	χ ²	df	p	
Male	38	69%	30	75%	66	66%	1.08	1	0.582	
Racial characteristics							41.23	2	<0.001	
Caucasian	43	78%	9	23%	54	54%				
African American	10	18%	30	77%	34	34%				
Other	2	4%	0	0%	12	12%				
Hispanic ethnicity	2	4%	0	0%	21	21%	15.57	1	<0.001	
Informant Type							24.31	1	<0.001	
Clinician	22	40%	0	0%	17	17%				
Friend or Relative	33	60%	40	100%	83	83%				
Residential Status							78.99	3	<0.001	
Independent and financially responsible	7	13%	24	60%	57	57%				
Independent, not financially responsible	11	20%	5	13%	17	17%				
Unsupervised residential facility	4	7%	5	13%	23	23%				
Supervised residential facility	33	60%	6	15%	3	3%				
Employment Status							10.12	3	0.120	
Employed part-time	10	18%	3	8%	7	7%				
Employed full-time	3	5%	2	5%	1	1%				
Unemployed	40	73%	35	88%	89	89%				
Retired	2	4%	0	0%	3	3%				
	Mean	SD	Mean	SD	Mean	SD	F	df	p	
Age	35.78	14.13	47.3	8.58	47.25	8.89	23.27	2, 192	<0.001	
Years of Education	14.13	2.81	13.07	1.64	12.32	2.34	9.95	2, 178	<0.001	
Baseline Scores										
Positive and Negative Syndrome Scale (range: 30–210)	62.11	13.55	64.55	15.86	62.21	14.47	0.43	2, 192	0.653	
UCSD Performance-Based Skills Assessment–Brief Version (range: 0–100)	75.47	14.38	81.65	13.19	75.32	12.02	3.73	2, 192	0.026	

Variable	Site										Analysis		
	Skyland Trail (n=55)			Atlanta VA Medical Center (n=40)			UCSD (n=100)						
	n	%		n	%		n	%		χ^2	df	p	
Everyday Functioning Battery (range: 0–13)	9.51	3.36		9.06	3.52		8.18	4.02		1.93	2, 168	0.148	
MATRICES Consensus Cognitive Battery, modified composite (average T score)	39.01	7.87		37.75	6.3		37.34	6.64		1.03	2, 191	0.358	

Table 2

Factor Loadings and P Values on the Ability Latent Trait in the Full Baseline Model, Based on Three Performance-Based Indicators and Examiner-Generated Total Scores for Six Real-World Functioning Scales^a

Variable	Analysis		
	Standardized factor loading	t	p
Performance-Based Ability Measures			
UCSD Performance-Based Skills Assessment–Brief Version	0.33	3.13	.005
Everyday Functioning Battery	0.55	4.14	.001
MATRICES Consensus Cognitive Battery, modified composite	0.57	4.26	.001
	Nonstandardized factor loading	t	p
Ratings of Real World Functioning			
Life Skills Profile	4.459	3.116	.152
Social Behavior Schedule	1.761	0.617	.537
Social Functioning Scale	1.451	0.304	.192
Independent Living Skills Survey	1.683	0.758	.448
Quality of Life Scale	1.206	1.271	.204
Specific Levels of Functioning	7.093	2.435	.015

^aOverall model Fit: $\chi^2 = 78.100$, degrees of freedom (df) = 56, p-value (p) = .027, and root mean square error of approximation = .078, with a 90%-confidence interval (.027, .117). Total variance shared between the ability latent trait and all 6 real-world measures: $R^2 = .41$.