Schizophrenia Research 210 (2019) 30–38

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

Comprehensive review of the research employing the schizophrenia cognition rating scale (SCoRS)

Philip D. Harvey [a,b], Anzalee Khan [c,d], Alexandra Atkins [c], Trina M. Walker [c,e], Richard S.E. Keefe [c,e,*]

[a] Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, FL, United States of America
[b] Research Service, Miami VA Healthcare System, Miami, FL, United States of America
[c] VeraSci, Inc., Durham, NC, United States of America
[d] Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, United States of America
[e] Department of Psychiatry, Duke University Medical Center, Durham, NC, United States of America

ARTICLE INFO

Article history:
Received 8 March 2019
Received in revised form 16 May 2019
Accepted 26 May 2019
Available online 10 June 2019

Keywords:
Schizophrenia
Cognition
Functioning
Co-primary
SCoRS
Rating scales

ABSTRACT

This review of research utilizing the Schizophrenia Cognition Rating Scale (SCoRS) outlines the development, evaluation, validation, and implementation of the SCoRS to assess whether the scale meets the criteria as a functional co-primary as defined by the MATRICS-CT initiative. Interview-based co-primary assessments should be: 1) practical and easy to administer for a clinician or researcher; 2) validated in individuals with schizophrenia; 3) contain the relevant areas of cognition and functioning applicable to schizophrenia; 4) able to assess all phases and severity levels of schizophrenia; 5) capable of monitoring disease progression; 6) minimal burden to patients; and 7) sensitive to assess treatment effects. A review of the literature was conducted to present information on the development, psychometric properties and usage of the SCoRS. Review of the development of the SCoRS followed the parameters outlined for scale development on content expert validation and feedback. The SCoRS shows good psychometric properties in various studies, and demonstrates low burden on clinicians and patients. The items measure global concepts that do not require notable cultural modification, making international use feasible. While multiple performance-based tests in cognition and functional outcomes are available, there is a need for a multi-domain, interview-based assessment of cognitive progression and treatment response in clinical trials. The SCoRS appears to meet many of the criteria for an optimal co-primary measure for schizophrenia cognition clinical trials as defined in the MATRICS-CT initiative.

© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Cognitive impairments are common in people with schizophrenia spectrum conditions and are associated with a variety of impairments in everyday functioning, including vocational, activities, social outcomes, and overall quality of life (QOL). In fact, deficits in cognition more significantly erode a patient's quality of life than any other aspect of the disorder, including psychosis. As a result, improvements in function and quality of life represent the ultimate treatment goal, and improvement of cognition has emerged as the direct target of treatment for many clinical trials. Treatment of cognitive impairments with either pharmacological or computerized cognitive training (CCT) strategies has been endorsed through consensus of industry, academic, and regulatory constituencies. This methodology has several critical features: clinical stability of participants, assessment of cognitive outcomes, and use of co-primary measures to index the potential clinical relevance of treatment-related cognitive changes. These co-primary measures can be performance-based measures of functional capacity or interview-based measures of cognitive and related everyday functional outcomes (Buchanan et al., 2005; Green et al., 2008). Several different interview-based measures have been developed; this review focuses on one of them, the Schizophrenia Cognition Rating Scale (SCoRS). We describe the development, evaluation, validation, and implementation of the SCoRS, and evaluate the instrument on its ease of administration, cross-cultural applicability, and sensitivity to treatment.

2. Criteria for a suitable co-primary measure

Food and Drug Administration (FDA) has guided drug developers focused on cognitive improvement to demonstrate not only improvement on cognitive endpoints, but also improvement on a measure of functional capacity (Buchanan et al., 2005). In people with schizophrenia, cognitive and functional impairments are present, so a drug will not be considered effective without demonstration of a treatment effect on both. Multiple primary endpoints are commonly designated as co-primary when it is necessary to demonstrate an effect on each of the

* Corresponding author at: Department of Psychiatry, Duke University Medical Center, DUMC Box 3270, Durham, NC 27710, United States of America.
E-mail address: richard.keefe@duke.edu (R.S.E. Keefe).

https://doi.org/10.1016/j.schres.2019.05.040
0920-9964/© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

endpoints (U.S. Department of Health and Human Services Food and Drug Administration, January, 2017). As defined in the MATRICS Coprimary and Translation (MATRICS-CT) initiative (Green et al., 2011), a co-primary measure must possess all the psychometric characteristics required for a primary cognitive performance measure (inter-rater and test retest reliability, and utility as a repeated measure) as well as convergence with cognitive performance on the one hand and real-world functioning on the other. Further requirements include ease of administration for testers (set up, scoring, lack of missing data), tolerability for patients, and limited duration of assessment. Sensitivity to treatment was not recommended as a requirement because no previous treatments had demonstrated clear success. While the SCoRS was included in the initial MATRICS co-primary validation study (Green et al., 2008), it was not included in the MATRICS-CT initiative because of analyses that had suggested that it overlapped with another scale, the Cognitive Assessment Interview (CAI). However, subsequently the CAI was viewed by those designing clinical trials for cognitive impairment in schizophrenia to have significant burden for assessors and participants; thus, the SCoRS has been widely used in many different clinical trials (e.g., Harvey et al., 2011; Keefe et al., 2015a; Brown et al., 2018). In this paper, we will review the data collected on the SCoRS.

3. Information sources and search strategy

We searched electronic databases which include Medline (2006–December 2018), PsychInfo (2006–December 2018), and PubMed (2006–December 2018) using key words "SCoRS, Schizophrenia Cognition Rating Scale," and reference lists from identified studies and review articles. We identified 310 articles, out of which there were 18 publications from independent studies that used the SCoRS in some way and provided information allowing for examination of psychometric and treatment response characteristics.

4. Assessment options for functional co-primary measures

While a "co-primary" functional endpoints aims at indexing a clinically meaningful benefit to the patient, there is an open question regarding which forms of functional assessment are the most valid and useful. Performance-based measures of functional capacity, which rely on patient task performance in a clinical setting, demonstrate close associations with cognitive performance and have the advantage of reducing the bias associated with some interview-based strategies (Harvey et al., 2007). Performance-based measures such as the University of California, San Diego Performance-Based Skills Assessment (UPSA) and Virtual Reality Functional Capacity Assessment Tool (VRFCAT), for instance, have been adopted in many trials due to strong psychometrics, including excellent test-retest reliability and strong correlations with cognitive performance (Harvey et al., 2007; Heinrichs et al., 2006; Mausbach et al., 2007; Patterson et al., 2001; Keefe et al., 2016). The UPSA is a performance-based measure of functional abilities with ecological validity; it consists of equivalently weighted measures of real-world activities (for example, planning an activity, such as a trip to the zoo, determining a bus or trolley route, dialing a phone number, selecting appropriate medications, and writing a check). The UPSA was validated in middle-aged and older healthy adults and in an outpatient sample of individuals with schizophrenia (Mausbach et al., 2007; Patterson et al., 2001), but has been employed across the age-span (Vesterager et al., 2012) and across other diagnoses (See Harvey et al., 2017, for a review). The VRFCAT is a computerized test of a patient's ability to navigate everyday situations. The VRFCAT demonstrated excellent test-retest reliability, strong discrimination between individuals with schizophrenia and healthy controls, and reduced practice effects in comparison to the UPSA (Keefe et al., 2016). It has also been used in other conditions (Atkins et al., 2015, 2018).

When appropriately implemented, there are many advantages to performance-based assessments. For example, performance-based measures are free of the potential informant/caregiver biases or inadequate knowledge of the individual's everyday functioning that may misrepresent informant-based accounts. However, all performance-based assessments require adaptation and validation before use across cultures, meaning that there may be occasional barriers to use. As technology and functional demands are constantly changing, performance-based assessments need to be easily modifiable and this requires constant attention. Further, in clinical settings, it may be easier to collect interview-based information.

5. Advantages of interview-based assessments

There are also advantages of interview-based methods for the assessment of cognitive functioning and related outcomes. First, interviews can be targeted toward detailed assessment of the performance of everyday tasks in the real world; performance-based measures, regardless of their veridicality, are assessments of task performance in a simulation. As a result, performance-based measures do not provide information about the ***likelihood*** of performing these skills in the real world, including initiation or persistence of efforts. It is noteworthy that the SCoRS does provide this type of information about functioning. Interview-based methods have flexibility and can be updated rapidly in the face of changes in technology or other common activities. For instance, when performance-based assessments are updated the process requires psychometric validation of new tasks. In the case of interview-based assessments, if the questions are written in a generalized manner, they can be immediately updated to include newly introduced tasks. For example, the SCoRS includes a question about managing money which gives general examples that are not restricted to a specific strategy or technology: making change and paying bills. Thus, the method of bill payment, using cash, writing a check, or paying online or on a hand-held device, is not specified and can be queried flexibly. In contrast, requiring someone to write a check or make a paper bank deposit, which many people no longer do, may lead to performance failures simply on the basis of lack of lifetime experience. Similarly, capability to utilize newly developed technology, ranging from household objects to computers, can be queried directly within the existing interview structure.

Another advantage of interview-based assessments is the opportunity to index the perspective of the patient in terms of their cognitive and functional abilities. People with schizophrenia have limitations in their awareness of illness in several different domains, including clinical symptoms (Markova and Berrios, 1992), cognitive impairments (Gould et al., 2015), social cognitive limitations (Silberstein et al., 2018), everyday disability (Durand et al., 2015), meta-cognitive management of task performance (Cornacchio et al., 2017), emotional state (Ouzir et al., 2012), and even tardive dyskinesia (Strassnig et al., 2018). Only about 50% of people with schizophrenia who manifest objective NP deficits endorse cognitive deficits and, in some studies there is no overall correlation between performance on NP tests and domain specific subjective impairments (Keefe et al., 2006). Further, in some cases, patients who report *more* cognitive functioning impairments perform *better* on measures of meta-cognitive self-assessment than patients who report no impairments. Thus, ability of people with schizophrenia to make a general assessment of their cognitive abilities in an interview context appears limited. An easily implemented strategy to examine awareness of cognitive limitations with interview-based assessment is to ask patients and informants the same questions about the patient's functioning. Discrepancies between reports can index both accuracy of self-assessment and directional biases, such as the tendency to over- or underestimate performance relative to informant reports and objective information. This assessment has important implications because overestimation of ability has been linked to impairments in real-world functional outcomes (Gould et al., 2015; Silberstein and Harvey,

2019). The SCoRS is designed to ask similar questions to informants and subjects, with easily calculated difference scores for individual items and global impressions.

Despite the advantages of interview-based methods, there are measurement issues that can affect test sensitivity. One of these is the need to interview informants. This strategy can be affected by several different biases, including halo effects, omission of key information, or gaps in knowledge of the individual, which can result in either over-estimation or under-estimation of an individual's proficiency (Loewenstein et al., 2001). One way to increase sensitivity in these interview-based measures might be to use performance-based measures, including performance on neuropsychological tests, as a reference point for the quality of the reports of different classes of informants (Sabbag et al., 2011).

6. Previous interview-based strategies targeting cognition

Interviews conducted with individuals with schizophrenia have been examined in several different studies for their convergence with performance-based assessments of cognition. Several studies have reported on the correlations between clinical ratings on the Positive and Negative Syndrome Scale (PANSS) Cognitive Factor and performance on neuropsychological (NP) tests (Bell et al., 1994; Ehmann et al., 2004; Hofer et al., 2007; Nielsen et al., 2014; Rodriguez-Jimenez, et al., 2013). The PANSS Cognitive Factor is composed of several PANSS items that vary somewhat across different factor analytic studies. Bell et al. (1994) assessed the validity of the cognitive factor in 147 patients diagnosed with schizophrenia or schizoaffective disorder and found a significant negative correlation between cognitive factor scores and performance on all the NP tests. Other studies find relatively limited correlations with NP test performance (Hofer et al., 2007). Overall, correlations between the PANSS Cognitive Factor and NP tests have been reported to be low to moderate, ranging from 0.20 to 0.53 (Bryson et al., 1999; Cameron et al., 2002; Daban et al., 2002; Hofer et al., 2007), which is considerably less than the typical correlations between performance -based assessment of functional capacity and cognitive performance. Additional studies have shown that correlations between NP test performance and the PANSS Cognitive Factor were smaller than those between NP performance and negative symptoms in studies of both chronic schizophrenia (Harvey et al., 2001) and drug naïve first episode patients (Good et al., 2004), suggesting that the PANSS Cognition Factor lacks specificity.

7. Development Strategy for the SCoRS

The following sections outline the steps of scale development undertaken for the SCoRS.

(1) Domain identification and item generation,
(2) Content expert validation, and
(3) Pilot testing.

8. Domain identification and item generation

From the outset, the SCoRS was developed to assess domains of cognition that were functionally relevant and observable. In this stage, the primary concern was content validity. A standard rational-empirical approach was adopted wherein the opinions of assessment experts were considered during the early development process. To generate themes and obtain more substantive insights pertinent to expertise, a panel was formed, composed of individuals with assessment expertise.

Eighteen different questions were developed, spanning functionally relevant activities that require different cognitive abilities including attention, memory, reasoning and problem solving, working memory, language production, and processing speed. Items are scored on a 4-point (1–4) scale with 1 reflecting no problems and 4 the most severe problems. A global rating ranging from 1 to 10 is also assigned and the follow-up version of the scale has questions about improvement or worsening since treatment began.

An initial item pool was developed based upon the Brief Cognitive Scale for patients with Mild Cognitive Impairment (Krishnan et al., 2001), and edited after pilot testing with a small group of people with schizophrenia. All pilot interviews included an interview with the patient and an informant, as well as generation of interviewer ratings based on the two interviews. Items were edited, a final item pool was developed, and an initial reliability and validity study was performed.

9. Content expert validation

The expert panel generated relevant items that were then subjected to content expert validation to ensure the content validity of the measurement instrument. The content experts were asked to address three elements in examining the SCoRS: representativeness, comprehensiveness, and clarity, to ensure that there were no ambiguous and poorly written items.

Another validation strategy was to measure concurrent validity with NP test performance as well as performance-based measures of functional capacity. Further, separate interviews were conducted with informants and patients, as well as a consensus judgment generated by an interviewer. As noted above, this strategy collected separate informant reports from the first stages of development of this scale, allowing both for assessment of the convergence between reports from different sources of information and the generation of a "consensus" judgment based on multiple, clearly identifiable, information sources.

10. Initial validity study (Keefe et al., 2006)

10.1. Step 1: A pilot study of the items of the measure was conducted

Sixty patients with DSM-IV schizophrenia were assessed with a battery of cognitive and functional measures, including the SCoRS, with 3 not providing complete data. Forty-seven (78%) of the patients who provided data were male. All patients were receiving antipsychotic medications.

Complete administration of the SCoRS included an interview with the patient, an interview with an informant for the patient and a rating by the interviewer who administered the scale to the patient and informant. Informal time estimates suggest that each interview required an average of about 12 min of interview time and one or 2 min of scoring time. The informant based his or her responses on interaction with and knowledge of the patient; we aimed to identify the informant as the person who had the most regular contact with the patient in everyday situations. In this study, all of the informants were staff members. The interviewer's rating reflected a best estimate combination of the two interviews incorporating the interviewer's observations. The informant ratings were completed within 7 days of the administration of the patient interview.

10.2. Step 2: Reliability assessment

10.2.1. Inter-rater reliability

Two interviewers participated in the same interview of 11 patients. Intra-class correlations of the relationship between the ratings generated by the two interviewers were calculated to assess the inter-rater reliability of the 18 SCoRS items. Thirteen of the 18 items had an ICC of 1.00, indicating absolute agreement between the two interviewers for all 11 patients. The ICCs for 4 of the other 5 items were greater than 0.90, and the lowest ICC for an item ("do you have difficulty walking as fast you would like") was 0.81.

10.2.2. Internal consistency

Reporting of internal consistency reliability is a necessary part of the scale development process (Hinkin, 1995). To assess the reliability of an instrument based on internal consistency, the minimum level of Cronbach's coefficient alpha (Price and Mueller, 1986) is 0.70 for basic research measures, following Nunnally's (1978) suggestion. Internal consistency for the SCoRS was calculated using Cronbach's coefficient alpha, which was found to be 0.79. Examination of the contribution of the individual items to the total scale scores indicated that there were no items whose deletion would have improved the overall internal consistency of the scale more than 0.01.

10.3. Step 3. Validity

10.3.1. Construct validation

In order to examine the convergent validity of the SCoRS, the relationship of the SCoRS to measures of cognitive functioning, performance-based assessments of functional capacity and functional outcome in schizophrenia was examined. These assessments are described below, as are the analyses performed in order to determine the estimated validity of the SCoRS.

Cognitive performance was assessed with the Brief Assessment of Cognition in Schizophrenia (BACS) (Keefe et al., 2004). The BACS requires approximately 30 min and is devised for easy administration and scoring by non-psychologists. The battery of tests in the BACS includes brief assessments of reasoning and problem solving, verbal fluency, attention, verbal memory, working memory and motor speed. A composite score is calculated by comparing each patient's performance on each measure to the performance of a healthy control sample and averaging the standardized z-scores.

Everyday functioning was assessed with the Independent Living Skills Inventory (ILSI) (Menditto et al., 1999). The ILSI is a standard functional assessment instrument measuring the extent to which individuals can competently perform a broad range of skills important for successful community living. Each of the items is rated according to the extent to which an individual is able to perform a skill, as well as the extent of assistance or guidance required. An overall score was created for the ILSI by calculating a mean of all 11 subscales.

Functional Capacity was assessed with the UCSD Performance-Based Skills Assessment (UPSA) (Patterson et al., 2001). The UPSA assesses skills necessary for functioning in the community by asking patients to perform relevant tasks in five areas: household chores, communication, finance, transportation, and planning recreational activities. An overall score was calculated for the UPSA by taking a mean of the scores in each area.

In order to keep knowledge of a patient's objective cognitive performance from biasing SCoRS and ILSI ratings, the tester who assessed cognitive performance with the BACS and functional capacity with the UPSA was always a different person from the interviewer who assessed cognition with the SCoRS and real-world functioning with the ILSI. In addition, raters remained blind to the results of the measures they did not administer.

Interviewer and informant global ratings on the SCoRS correlated significantly with all the validators: cognitive performance, functional capacity, and everyday functioning. Patient-reported global ratings were not significantly correlated with any of the validators. For none of the validators did the individual patient or informant global ratings account for significant variance beyond that accounted for by the global interviewer rating. Finally, the SCoRS interviewer global rating predicted unique variance in real-world functioning as measured by the ILSI above and beyond the that predicted by the BACS and UPSA. When the order of the regression was reversed, the BACS and UPSA accounted for no variance in everyday functioning above and beyond the interviewer global ratings.

10.3.2. Conclusions from the initial validation study

The findings of the first study of the SCoRS suggested that interviewer judgments informed by interviews with both an informant and the patient are associated with everyday functioning as well as with cognitive performance. As a result, the SCoRS was found very early on to meet criteria for a suitable co-primary measure for cognition treatment trials. While the SCoRS global interviewer rating provided incremental information regarding everyday functioning beyond that NP performance and UPSA scores, the latter two variables did not improve on the prediction of everyday functioning past the SCoRS.

Findings from this study generated several suggestions regarding implementation of the SCoRS. Informant ratings are critical and are themselves highly correlated with everyday functioning, NP test performance, and functional capacity. Self-reported SCoRS ratings had no overlap with any of the validating measures and consequently suggest extreme caution in interpreting any self-report data. Results from treatment studies to be reviewed below suggest that this concern also applies to treatment outcomes.

Following the initial validation study, the SCoRS was evaluated by the MATRICS project as one of 4 potential co-primary measures for schizophrenia cognition trials. The MATRICS group eliminated one of the items assessing motor function and added 3 items assessing social cognition so that the domains to be assessed would be consistent with the cognitive domains in the MATRICS cognitive battery. The revisions resulted in a 20-item scale. Following an extensive psychometric study on the 20-item version, the SCoRS had very similar characteristics to the 18-item version in the original validation study (Keefe et al., 2006). The MATRICS group concluded that the SCoRS had good tolerability for patients and informants, good practicality for raters, medium sized (r~0.30) correlations with functioning and cognition, very good psychometric characteristics, and very few missing data (Green et al., 2008). Subsequent work with the SCoRS used the 20-item version.

11. Replication of findings

Additional validity data on the SCoRS was obtained in a subsequent study (Keefe et al., 2016) where patients were examined with the MCCB and the UPSA, and informant and patient reports were collected with the SCoRS. In addition, ratings of everyday functional outcomes were collected with the Specific Levels of Functioning (SLOF) (Schneider and Struening, 1983), a rating scale that interviews patients' informants to assess patients' everyday functioning in domains of interpersonal functioning, everyday activities, and work skills.

There are several important features of this study that have not been published previously. The correlation of MCCB performance and SCoRS ratings in a completely independent sample provides additional validity data for the SCoRS as a correlate of cognitive performance. Similarly, correlations of NP test performance versus SCoRS ratings with functional outcomes gives additional information about the usefulness of the SCoRS as a co-primary measure validated against RW outcomes. Finally, comparisons of the UPSA and the SCoRS for prediction of everyday outcomes provide information about the differential validity of these two measures of functional capacity.

Table 1 presents the scores on all of the variables of interest and Table 2 presents the intercorrelations of those variables. Using $P < .05$ as a threshold, MCCB composite scores were correlated with the UPSA, both SCoRS ratings (informant and interviewer), as well as SLOF work, everyday activities, and composite scores but not interpersonal functioning. The UPSA was correlated with the same variables, other than SCoRS informant ratings and everyday activities. SCoRS interviewer ratings and informant ratings were correlated with all variables: the UPSA, the MCCB, and all 4 SLOF variables.

As the pattern of correlations for the MCCB and UPSA were similar, predicting work and composite SLOF scores, we performed comparative regression analyses on the SLOF global scores only. For the first analysis, we used a stepwise regression analysis to predict SLOF Global scores

Table 1
Scores on cognitive performance, SCoRS ratings, performance-based functional capacity and everyday functioning.

	M	SD	Metric	Range
UPSA Total Score	70.85	11.80	Score	34–96
SCoRS Variables				
Informant Total Score	35.81	9.59	Score	20–62
Interviewer Total Score	38.82	9.90	Score	21–61
MCCB Neurocognitive composite	27.87	12.83	T-score	1–65
SLOF Subscales				
Interpersonal	22.95	5.51	Score	11–35
Everyday Activities	30.87	4.22	Score	17–45
Work	20.67	4.86	Score	8–30
Global	74.49	11.02	Score	36–100

Notes. UPSA: UCSD Performance Based Skills Assessment.
SCoRS: Schizophrenia Cognition Rating Scale.
MCCB: MATRICS Consensus Cognitive Battery.
SLOF: Specific Levels of Functioning.
All Variables other SCoRS Informant, $n = 159$; SCoRS Informant Ratings, $n = 77$.
From: Keefe et al., 2016.

with UPSA, MCCB, and SCoRS interviewer rated scores. For the second analysis, we substituted the interviewer ratings on the SCoRS with the informant ratings. For the first regression, ($F(1,157) = 60.29$, $p < .001$) only interviewer SCoRS ratings entered the equation, accounting for 28% of the variance ($t = 7.78$, $p < .001$) in SLOF composite scores. Neither the MCCB, $t = 1.0$, $p = .32$, nor the UPSA, $t = 0.79$, $p = .43$, entered the equation. For the second analysis, using SCoRS informant ratings with a smaller sample size of $N = 77$ ($F(1,75) = 17.62$, $p < .001$), again the only variable that entered was the SCoRS informant ratings, accounting for 19% of the variance ($t = 4{,}20$, $p < .001$). Neither the MCCB ($t = 0.29$, $p = .77$) nor the UPSA ($t = 0.14$, $p = .89$) entered the equation.

Final analyses predicted functioning as measured by the SLOF by first entering the UPSA and MCCB into the equation as a block and then entering the SCoRS ratings. The overall analysis ($F(3,155) = 20.31$, $p<001$) was significant and suggested that the MCCB and UPSA accounted for 8% of the variance in the first block ($t = 2.7$, $p = .007$) while the SCoRS Rating generated by the interviewer entered in the second block and accounted for 20% of the variance, ($t = 6.54$, $p < .001$). When the analysis was repeated for the informants SCoRS ratings, the results were similar. In the overall analysis ($F(3,73) = 5.75$, $p < .001$), the MCCB and UPSA did not enter the equation in the first block ($t = 1.35$, $p = .18$), but the informant SCoRS did enter in the second block, accounting for 16% of the variance ($t = 3.75$, $p < .001$).

These analyses suggest that SCoRS ratings meet all criteria for a suitable co-primary measure, in that the SCoRS is correlated with both cognitive performance and with real-world outcomes. In all cases, SCoRS ratings were more strongly correlated with all of the SLOF-measured functional outcomes than the UPSA. The correlations between interviewer SCoRS and SLOF ratings should be viewed in the context of overlap of raters and method variance. Critically, the informant ratings on the SCoRS are not the origin of the SLOF ratings, which are generated by the interviewer, yet the correlations between the informant SCoRS ratings and the interviewer ratings on the SLOF were very similar to those that were generated by the same rater. In addition, the informant SCoRS ratings have excellent evidence of convergence with the MCCB, sharing 13% of the variance with those performance-based data about which the informant and interviewer had no information.

12. Additional reliability analyses

To gather more test-retest reliability data on the SCoRS, a set of analyses based on data taken from a large clinical trial was conducted, with these results previously published in Keefe et al. (2015a). All patients self-reported their functioning, most of the patients also had informant ratings, and an examiner rating was generated. 319 patients with DSM-IV schizophrenia were randomized at 32 sites in the United States, Russia, Ukraine, and Serbia in the context of a phase 2 clinical trial to determine the safety and efficacy of encenicline, an alpha-7 nicotinic agonist, for the treatment of cognitive impairment in patients with schizophrenia. Considering Day 0 as the day that treatment was initiated, the SCoRS was administered in this study at Day −14, Day −7, twice on Day −4, and then on post-treatment visits at Day 28, Day 56 and Day 77.

12.1. Pre-treatment - all participants

The ICCs for test-retest reliability were greater than 0.90 for the patient ($N = 299$), informant ($N = 231$) and interviewer ($N = 302$) SCoRS total scores and the separately generated interviewer global rating. When interviewer and patient ICCs were re-examined using only assessments where an informant was available ($N = 231$), both remained above 0.90. ICCs for the pre-treatment period were also compared between sites in the US (when an informant was present, $N = 90$) and Europe (informants present at all assessments, $N = 141$). ICCs were higher in Europe than the US for the patient (0.95 vs. 0.86), informant (0.97 vs. 0.82), and interviewer (0.97 vs. 0.84) SCoRS total, and interviewer global rating (0.99 vs. 0.86). While the Ukrainian ICC values were surprisingly high, all of the US ICC values were excellent, and would certainly not be too low for use as a clinical outcomes measure.

The ICCs for individual SCoRS items varied by question and source of information. In general, the ICC values were lowest for the patient (19 of 20 items) and highest for the interviewer (12 of 20 items). Overall, the mean values (range) were 0.77 (0.69–0.82) for patient, 0.81 (0.74–0.88) for informant, and 0.82 (0.71–0.87) for interviewer.

Table 2
Intercorrelations of cognitive performance, SCoRS ratings, performance-based functional capacity, and everyday functioning.

	1	2	3	4	5	6	7	8
1. UPSA Total	1.0	−0.28**	−0.22	0.70***	0.08	0.14	0.27***	0.21**
2. Clinical SCoRS		1.0	0.89***	−0.44***	−0.36***	−0.43***	−0.44***	−0.53***
3. Informant SCoRS			1.0	−0.37***	−0.26*	−0.48***	−0.30**	−0.44***
4. MCCB Composite				1.0	0.13	0.21*	0.39***	0.31***
5. SLOF Interpersonal					1.0	0.26***	0.46***	0.79***
6. SLOF Activities						1.0	0.45**	0.70***
7. SLOF Work							1.0	0.82***
8. SLOF Global								1.0

Notes. UPSA: UCSD Performance Based Skills Assessment.
SCoRS: Schizophrenia Cognition Rating Scale.
MCCB: MATRICS Consensus Cognitive Battery.
SLOF: Specific Levels of Functioning.
All Variables other SCoRS Informant, n = 159; SCoRS Informant Ratings, n = 77.
From: Keefe et al., 2016.
* $P < .05$.
** $p < .01$.
*** $p < .001$.

12.2. Post-treatment - placebo group

Assessment of the test-retest reliability of the SCoRS was also examined for the treatment period in the placebo group only. The ICCs for the SCoRS total were 0.76 for the patient ratings ($N = 100$), 0.85 for the informant ratings ($N = 71$) and 0.81 for the interviewer ratings ($N = 101$). The ICC for the interviewer global rating was somewhat lower than for the SCoRS total (0.76 vs. 0.81). Again, individual items varied considerably, and ICCs were substantially lower than for the pre-treatment comparisons. The mean (range) ICC was 0.68 (0.55–0.78) for the patient ratings, 0.71 (0.59–0.80) for the informant ratings, and 0.72 (0.47–0.83) for the interviewer. These data suggest considerable stability of scores, particularly during pre-treatment periods. There was clearly some increased error associated with re-ratings during a period when the ratings may have been biased by patients and raters expecting a treatment effect.

13. Exploratory factor analysis of the SCoRS

Another study of 76 patients with DSM-IV schizophrenia assessed in the context of a multi-center validation study of functional capacity outcome measures, including the SCoRS, was conducted (Atkins et al., 2017). All patients had informant data. Assessments were completed at two visits 7–14 days apart by highly trained raters at the University of Miami Medical Center, University of California at San Diego, and the University of South Carolina. The data from this study were combined with the 231 patients with informant ratings described above.

An exploratory factor analysis was run on the 20 interviewer items of the SCoRS using the first ratings performed from the combined dataset. Missing items led to the loss of 5 observations. The objective of the exploratory factor analysis was to identify those items that most clearly represent the content domain of the underlying construct. There are no hard and fast rules for this, but the 0.40 criterion level appears most commonly used in judging factor loadings as meaningful (Ford et al., 1986), as it is consistent with the correlations expected between independent indicators of a scientific construct in classical measurement theory (Cronbach and Meehl, 1955). The exploratory factor analysis on the combined dataset ($N = 295$) indicated that a single factor was the best structure. The factor loadings were used as weights to create a factor score. When that unifactorial score was compared to the total score as a treatment outcomes measure, the results were identical for total scores and factor scores.

14. Evolving uses of the SCoRS

The SCoRS has been translated and validated in several different languages and cultures and has been administered in patients at risk for psychosis (Higuchi et al., 2017). Finally, as described below, it has been used as a treatment outcomes measure in several different completed clinical trials, including multinational trials where the international validated measures were used.

15. International validation of the SCoRS

Versions of the SCoRS have been developed in 22 languages and deployed in multiple countries (see Table 3). Published studies on the international validation of the SCoRS span several different countries: Singapore, Japan, Iran, France, Brazil, and Italy. In these studies, cognitive performance was found to correlate with the interviewer and informant scores (when examined) similarly to studies in the US, although some differences were reported due to various patient population differences (Mazhari et al., 2017; Higuchi et al., 2017; Vita et al., 2013; Chia et al., 2010). In all studies where examined, SCoRS total scores were also significantly correlated with everyday functioning. Further, in a qualitative research study, (Gonzalez et al., 2013) it was also reported that interview-based measures such as the SCoRS are most suitable for use in international treatment trials, because the measures require considerably less cultural adaptation. That same study reported that some of the current set of performance based functional capacity measures (e.g., UPSA) required the performance of tasks that are unfamiliar to people in several different countries, particularly in rural areas. The SCoRS was judged to not assess performance of functional tasks that are differentially likely to be performed in different countries.

Table 3
Availability of professional and academic versions of the SCoRS in international languages.

SCoRS		
Language	Professional translation	Academic
Bulgarian	X	
English	N/A	
Chinese (Simplified)	X	
Chinese (Traditional)	X	
Czech	X	
Dutch (Netherlands)	X	
French (CA)	X	
French (FR)	X	
German	X	
Greek		X
Iranian		X
Italian	X	
Japanese	X	
Korean	X	
Polish	X	
Portuguese (Brazil)		X
Romanian	X	
Russian	X	
Serbian		X
Spanish	X	
Turkish		X
Ukrainian	X	

16. Use of the SCoRS across clinical state and different schizophrenia spectrum conditions

According to the MATRICS guidelines discussed above, cognitive enhancement clinical trials should be performed in clinically stable patients with a diagnosis of schizophrenia. However, in clinical use outside of formal clinical trials, it is possible that other participants might be evaluated with this assessment tool. Several studies have addressed the validity of the SCoRS in populations that would not qualify for a MATRICS-type of clinical trial at the current time.

In a comparative study in Japan of clinical high-risk patients ($n = 33$), schizophrenia patients ($n = 45$) and healthy controls ($n = 63$), the SCoRS was significantly correlated with everyday functioning for both schizophrenia and clinical high-risk patients. For the clinical high-risk patients, the SCoRS had a higher correlation with everyday functioning ($r = 0.79$) than for individuals with schizophrenia ($r = 0.45$) (Takahashia et al., 2017). In a similar study that also included patients with first episode psychosis (Higuchi et al., 2017), correlations between BACS cognitive performance and SCoRS global ratings was highest in first episode patients ($r = 0.50$), who performed worse on the BACS than the other groups, next highest in chronic schizophrenia patients ($r = 0.38$), and weakest in people at risk for psychosis ($r = 0.28$), while correlations between SCoRS global ratings and functioning measured with the SOFAS and SCoRS was similar in chronic patients ($r = 0.55$) and at-risk patients ($r = 0.51$) and lowest in first episode patients ($r = 0.37$). This intriguing set of results suggests that severity of cognition-related functioning is more strongly correlated with cognitive performance in early phases of illness, but later on in chronic patients, cognition-related functioning is related to overall level of day-to-day functional skills.

When comparing clinically stable and acutely psychotic patients on the validity of the SCoRS in a study in Italy, there were differences in patterns of correlation as a function of symptom severity (Vita et al., 2013).

In patients with total PANSS scores less than 72, the correlation of the SCoRS and NP performance was $r = 0.52$, $p < .001$. In patients with higher PANSS scores, the correlation was smaller at $r = 0.26$, suggesting that only 7% of the variance is shared between symptoms and report of cognitive difficulty. Thus, factors other than cognition are likely to be impacting on the behaviors measured by the SCoRS in patients with more severe psychosis. The contrasting results in acutely psychotic patients suggest that the SCoRS may be vulnerable to less stability over short periods of acute illness. They also raise the possibility that treatment responsiveness of the SCoRS in more highly psychotic patients could be attributable to symptom reduction in addition to primary cognitive impairment.

17. Treatment sensitivity of the SCoRS

Interview-based assessments of cognition and functioning have several features that lead to them to be promising candidates for treatment outcomes measures. First, as they are not performance-based, there should be no practice effects from repeated exposure to assessment materials. There is the possibility that informants could have changes in their sensitivity to patients' behavior and functional limitations by being interviewed about their functioning, so it may be desirable to have a pre-study exposure to the procedure of being an informant. It is not possible to predict for any individual or group of patients whether being interviewed would increase or decrease scores, although placebo effects with the SCoRS have been reported (Keefe et al., 2015b). Further, informants may also be susceptible to placebo effects, in that the optimistic expectations of functional gains by informants and raters could be reflected in changes in ratings due to their biases.

In the first study using the SCoRS as an outcome measure, Harvey et al. (2011) reported on a randomized short-term comparison of ziprasidone and lurasidone on neuropsychological assessment battery and the SCoRS. There was significant group x time interaction for neuropsychological test performance, and the lurasidone patients improved significantly from baseline. For the SCoRS, the results were similar in that the treatment difference was marginally significant, $p < .056$, with the lurasidone treated patients improving significantly from baseline while the ziprasidone patients did not.

Several other studies had negative results for both NP test performance and the SCoRS. For example, MK-0777, a gamma- aminobutyric acid (GABA)A _2/_3 partial agonist, was not found improve either NP performance or SCoRS total scores (Buchanan et al., 2011). Davunitide, a neuroactive peptide, was found to have marginally significant improvements in cognitive performance, significant improvements on the UPSA, and no changes on the SCoRS in a short-term trial (Javitt et al., 2012). However, there was a clear problem in the study in that the placebo group performed significantly better than the treatment group and was at ceiling at baseline. Another published pharmacological study examined change in cognitive performance with the SCoRS as a co-primary outcome (Brown et al., 2018). That study was negative for the effects of the pharmacological treatment on cognition and the SCoRS also did not respond to treatment. Similarly, Kane et al. (2010) found similar results with armodafinil versus placebo. Other studies on clinicaltrials.gov suggested that there was similar convergence between SCoRS changes and changes in NP test results. As a result, the convergence between the extent of response on performance on NP tests and the SCoRS has been consistent across trials that were both positive for treatment with agent under investigation and negative in their results.

In a study in Japan examining the effects of augmenting group occupational therapy (GOT) with individualized occupational therapy (IOT) on NP performance and the SCoRS, there was a statistically significant improvement on composite NP performance for the 69 patients receiving IOT plus GOT, ES = 0.44 (Shimada et al., 2018). For the SCoRS, the effect size was moderate when comparing the two treatments' changes from baseline, d = 0.30, but the group x time interaction was non-significant, $p = .24$. It should be noted that GOT alone may be an active treatment, and there was no inactive control, so the failure of the treatments to separate does not mean that the improvements in the SCoRS in the GOT group alone are not an actual treatment effect.

Finally, a study comparing two doses of encenicline, an alpha-7 partial agonist to placebo generated a very interesting pattern of results (Keefe et al., 2015b). In this study, NP test performance was compared, as was the SCoRS across treatments. Not all of the patients at the treatment sites had an informant who was available to provide ratings, so it was possible to compare the treatment sensitivity of interviewer SCoRS ratings that were based solely on patient report to ratings informed by both informant and patient information. The distributions of the availability of ratings were: total patients ($N = 299$), and total informants ($n = 231$). Analyses were performed for both SCoRS total and item scores based on patient reports and interviewer ratings informed by patient plus informant reports. We also present here the informant reports as well for information purposes.

The effect size for 1.0 mg/day of encenicline vs. placebo was d = 0.37, $p < .01$ for interviewer ratings based on informants and patients, d = 0.57, $p < .001$ for informant reports alone and the effect size for self-reports was essentially zero. 7/20 individual items responded to treatment with interviewer ratings and 8/20 responded with informant ratings alone. Thus, self-reported cognitive performance was not sensitive to treatment response while SCoRS ratings based on informant information separated from placebo. It is of likely importance that cognitive performance also responded to treatment with the higher dose of encenicline in that study (Keefe et al., 2015b).

18. Ongoing studies

At the current time, there are 7 active trials in ongoing with the SCoRS as a co-primary outcomes measure.

(clinicaltrials.gov/ct2/results?cond=schizophrenia&term=SCoRS&cntry=&state=&city=&dist=).

These studies examine the effects of different medications, exercise, and psychosocial interventions on cognition in people with schizophrenia. Two other studies are apparently completed, with no results posted on the website or published to date.

19. Placebo effects

Some interview-based measures, particularly those targeted at clinical symptoms, have been noted to manifest increasing placebo effects over the course of the past two decades. Specifically, placebo effects with the PANSS in double-blind placebo-controlled trials examining the efficacy of antipsychotics have been increasing. As reviewed by Kemp et al. (2010), the average placebo effect on the PANSS increased from an average of a 3-point worsening in 1993 to a 12-point improvement around 2008. This increased placebo effect challenges the identification of a true treatment signal.

This issue is important to address for the detection of a cognitive improvement signal. Many schizophrenia cognition trials have been negative and the question of whether practice/placebo effects on the MCCB are responsible has been raised. To address this question, data from over 900 patients from 12 different clinical trials was analyzed to determine the magnitude of placebo/practice effects (Keefe et al., 2017). This analysis demonstrated that changes in performance on the MCCB that occurred during placebo periods did not differ from pre-randomization assessments where the participants knew that they were not receiving active treatment, thus ruling out a placebo effect. Further, the practice effect across retesting was very consistent and small, at a change of 0.1 SD per reassessment regardless of the number of reassessments and regardless of whether the study included a pre-randomization run-in period.

The SCoRS, as an interview-based measure should not have practice effects, but an examination of placebo effects is important. In the Javitt

et al. (2012) study, the SCoRS improved 0.05 SD at the first reassessment and 0.02 SD at the second reassessment in the placebo group and 0.26 SD at the first assessment and 0.40 SD at the second in the low-dose davunitide group. In the placebo group in that study, the MCCB improved by 0.30 SD in the placebo group, quite consistent with the SCoRS change and the Keefe et al., 2015a, overall results.

In another negative trial, Buchanan et al. (2011) reported that the MCCB improved by 0.11 SD in the placebo group and by a smaller amount in the two treatment groups. The SCoRS changed 0.09SD in the placebo group, again showing changes consistent with the MCCB. In a third negative trial using the SCoRS and MCCB (Brown et al., 2018), similar results were found. MCCB scores changed 0.25 SD over 2 reassessments in the placebo group while the SCoRS changed 0.17 SD over the same time period in this sample.

In a positive trial, where the drug separated from placebo on both the MCCB and the SCoRS, Keefe et al. (2015a) reported that the MCCB separated from placebo in the high dose group with an effect size of d = 0.28, while the SCoRS had an effect size of d = 0.36. Change from baseline in the placebo group was 0.20 SD over two assessments with the MCCB and 0.40 SD for the SCoRS over three reassessments. In a later reanalysis of the SCoRS data, focusing on changes associated with the type of informant, it was found that the patient-reported SCoRS ratings were much less likely to separate from placebo than informant or clinician ratings. Further, during placebo treatment, patient scores were more variable and less reliable than those generated by informants or interviewers.

20. Conclusions

The SCoRS appears to meet many of the criteria for an optimal co-primary measure for schizophrenia cognition clinical trials as defined in the MATRICS-CT initiative (Green et al., 2011). The SCoRS manifests consistently strong psychometric properties, including test-retest and inter-rater reliability, utility as a repeated measure, convergence with cognitive performance on the one hand and real-world functioning on the other. The SCoRS presents very low burden on patients, raters, and informants, and rapid completion of the assessments is the norm. The items measure global concepts that do not require notable cultural modification, making international use highly feasible.

Convergence between NP test performance and the SCoRS has been consistent across studies although it is consistently lower than correlations between NP tests and the UPSA. Ratings generated by clinician-interviewers and informants seem to have relatively similar properties while patient ratings seem less reliable and valid. Convergence between the SCoRS and everyday functioning has been consistently demonstrated and informant ratings are found to converge with clinician ratings of everyday functioning.

Consistent with considerable research evaluating self-assessment capability in schizophrenia, patient self-assessments appear to lack validity, to have more potential for placebo effects, and are less reliable. Patients who misestimate their cognitive performance also misestimate their everyday functioning, basing their estimates on other factors such as their mood state or if they are by themselves when asked to make a rating. Thus, there is very little valid treatment outcomes information that can be obtained, other than performance-based assessments, on patients who lack informants. As some studies suggest that as many of 25% of people with schizophrenia may lack a suitable informant, this is a limitation of instruments like the SCoRS that appear to require informant input to yield good reliability and validity. In such situations, valid performance-based assessments of functional capacity, particularly those with clearly similar alternative forms, may be a preferred strategy for assessing the clinical meaningfulness of a cognitive change in a clinical trial.

Sensitivity to treatment with the SCoRS has been demonstrated in multiple studies, including sensitivity to treatment effects also detected with NP tests. On the other hand, lack of change on the SCoRS has been associated with lack of improvements in NP test performance. Placebo effects seem to be minimal and quite consistent with placebo effects on the MCCB.

Funding

No funding source to report.

Contributors

Drs. Harvey, Khan, Atkins, and Keefe all participated in the drafting and reviewing of the manuscript. Analyses were performed by all of the authors. Ms. Walker was involved in rater training and data review for all of the data generated through Duke and VeraSci.

Declaration of Competing Interest

In the last three years, Dr. Harvey has received consulting fees or travel reimbursements from Allergan, Alkermes, Akili, Biogen, Boehringer Ingelheim, Forum Pharma, Genentech (Roche Pharma), Intra-Cellular Therapies, Jazz Pharma, Lundbeck Pharma, Minerva Pharma, Otsuka America (Otsuka Digital Health), Sanofi Pharma, Sunovion Pharma, Takeda Pharma, and Teva. He receives royalties from the Brief Assessment of Cognition in Schizophrenia and the MATRICS Consensus Battery. He is chief scientific officer of i-Function, Inc. He has a research grant from Takeda and from the Stanley Medical Research Foundation.

Dr. Anzalee Khan currently or in the past 3 years has received honoraria, served as a consultant, speaker, or advisory board member for National Institute of Mental Health (NIMH) and Teva Pharmaceuticals.

Dr. Alexandra Atkins currently or in the past 3 years has received honoraria, served as a consultant, speaker, or advisory board member for National Institute of Mental Health (NIMH) and National Institute of Aging (NIA). Dr. Atkins is a full-time employee of Verasci, Inc.

Ms. Walker currently or in the past 3 years has received an honorarium from Boehringer Ingelheim and receives compensation for her services from VeraSci.

Dr. Richard Keefe currently or in the past 3 years has received honoraria, served as a consultant, speaker, or advisory board member for Abbvie, Acadia, Aeglea, Akebia, Akili, Alkermes, Allergan, ArmaGen, Astellas, Avanir, AviNeuro/ChemRar, Axovant, Biogen, Boehringer-Ingelheim, Cerecor, CoMentis, Critical Path Institute, FORUM, Gammon Howard & Zeszotarski, Global Medical Education (GME), GW Pharmaceuticals, Intracellular Therapeutics, Janssen, Kempharm, Lundbeck, Lysogene, MedScape, Mentis Cura, Merck, Merrakris Therapetics, Minerva Neurosciences Inc., Mitsubishi, Montana State University, Monteris, Moscow Research Institute of Psychiatry, Neuralstem, Neuronix, Novartis, NY State Office of Mental Health, Orygen, Otsuka, Paradigm Testing, Percept Solutions, Pfizer, Pharm-Olam, Regenix Bio, Reviva, Roche, Sangamo, Sanofi, SOBI, Six Degrees Medical, Sunovion, Takeda, Targacept, Teague Rotenstreich Stanaland Fox & Holt, Thrombosis Research Institute, University of Moscow, University of Southern California, University of Texas Southwest Medical Center, WebMD, and Wilson Therapeutics. Dr. Keefe has currently or in the past 3 years received research funding from the National Institute of Mental Health and Boehringer-Ingelheim. Dr. Keefe receives royalties from versions of the BAC testing battery, the MATRICS Battery (BACS Symbol Coding), and the Virtual Reality Functional Capacity Assessment Tool (VRFCAT). He is also a shareholder in VeraSci, Inc. and Sengenix.

Acknowledgements

None.

References

Atkins, A.S., Stroescu, I., Spagnola, N.B., Davis, V.G., Patterson, T.D., Narasimhan, M., Harvey, P.D., Keefe, R.S., 2015. Assessment of age-related differences in functional capacity using the virtual reality functional capacity assessment tool (VRFCAT). J. Prev. Alzheim. Dis. 2 (2), 121–127.

Atkins, A.S., Tseng, T., Vaughn, A., Twamley, E.W., Harvey, P., Patterson, T., Narasimhan, M., Keefe, R.S., 2017. Validation of the tablet administered brief assessment of cognition (BAC-app). Schizophr. Res. 181 (Mar), 100–106.

Atkins, A.S., Khan, A., Ulshen, D., Vaughan, A., Balentin, D., Dickerson, H., Liharska, L.E., Plassman, B., Welsh-Bohmer, K., Keefe, R.S.E., 2018. Assessment of instrumental activities of daily living in older adults with subjective cognitive decline using the virtual reality functional capacity assessment tool (VRFCAT). J. Prev. Alzheimers. Dis. 5 (4), 216–234.

Bell, M.D., Lysaker, P.H., Milstein, R.M., Beam-Goulet, J.L., 1994. Concurrent validity of the cognitive component of schizophrenia: relationship of PANSS scores to neuropsychological assessments. Psychiatry Res. 54 (1), 51–58.

Brown, D., Nakagome, K., Cordes, J., Brenner, R., Grunder, G., Keefe, R.S.E., Riesenberg, R., Walling, D.P., Daniels, K., Wang, L., McGinniss, J., Sand, M., 2019. Evaluation of the efficacy, safety, and tolerability of BI 409306, a novel phosphodiesterase 9 inhibitor, in cognitive impairment in schizophrenia: a randomized, double-blind, placebo-controlled, phase II trial. Schizophr. Bull. 45 (2), 350–359.

Bryson, G., Bell, M., Greig, T., Kaplan, E., 1999. Internal consistency, temporal stability and neuropsychological correlates of three cognitive components of the positive and negative syndrome scale (PANSS). Schizophr. Res. 38 (1), 27–35.

Buchanan, R.W., Davis, M., Goff, D., Green, M.F., Keefe, R.S.E., Leon, A.C., Nuechterlein, K.H., Laughren, T., Levin, R., Stover, E., Fenton, W., Marder, S.R., 2005. A summary of the

FDA-NIMH-MATRICS workshop on clinical trial design for neurocognitive drugs for schizophrenia. Schizophr. Bull. 31 (1), 5–19.
Buchanan, R.W., Keefe, R.S., Lieberman, J.A., Barch, D.M., Csernansky, J.G., Goff, D.C., Gold, J.M., Green, M.F., Jarskog, L.F., Javitt, D.C., Kimhy, D., Draus, M.S., McEvoy, J.P., Mesholam-Gately, R.I., Seidman, L.J., Ball, M.P., McMahon, R.P., Kern, R.S., Robinson, J., Marder, S.R., 2011. A randomized clinical trial of MK-0777 for the treatment of cognitive impairments in people with schizophrenia. Biol. Psychiatry 69 (5), 442–449.
Cameron, A.M., Oram, J., Geffen, G.M., Kavanagh, D.J., McGrath, J.J., Geffen, L.B., 2002. Working memory correlates of three symptom clusters in schizophrenia. Psychiatry Res. 110 (1), 49–61.
Chia, M.Y., Chan, W.Y., Chua, K.Y., Lee, H., Lee, J., Lee, R., Lim, C., Tay, E., Woon, P.S., Keefe, R.S., Sim, K., 2010. The schizophrenia cognition rating scale: validation of an interview-based assessment of cognitive functioning in Asian patients with schizophrenia. Psychiatry Res. 178 (1), 33–38.
Cornacchio, D., Pinkham, A.E., Penn, D.L., Harvey, P.D., 2017. Self-assessment of social cognitive ability in individuals with schizophrenia: appraising task difficulty and allocation of effort. Schizophr. Res. 179 (Jan), 85–90.
Cronbach, L.J., Meehl, P.E., 1955. Construct validity in psychological tests. Psychol. Bull. 52 (4), 281–302.
Daban, C., Amado, I., Bayle, F., Gut, A., Willard, D., Bourdel, M.C., Loo, H., Olie, J.P., Millet, B., Krebs, M.O., Poirier, M.F., 2002. Correlation between clinical syndromes and neuropsychological tasks in unmedicated patients with recent onset schizophrenia. Psychiatry Res. 113 (1–2), 83–92.
Durand, D., Strassnig, M., Sabbag, S., Gould, F., Twamley, E.W., Patterson, T.L., Harvey, P.D., 2015. Factors influencing self-assessment of cognition and functioning in schizophrenia: implications for treatment studies. Eur. Neuropsychopharmacol. 25 (2), 185–191.
Ehmann, T.S., Khanbhai, I., Macewan, G.W., Smith, G.N., Honer, W.G., Flynn, S., Altman, S., 2004. Neuropsychological correlates of the PANSS cognitive factor. Psychopathology 37 (5), 253–258.
Ford, J.K., MacCallum, R.C., Tait, M., 1986. The application of exploratory factor analysis in applied psychology: a critical review and analysis. Pers. Psychol. 39 (2), 291–314.
Gonzalez, J.M., Rubin, M., Fredrick, M.M., Velligan, D.I., 2013. A qualitative assessment of cross-cultural adaptation of intermediate measures for schizophrenia in multisite international studies. Psychiatry Res. 206 (2–3), 166–172.
Good, K.P., Rabinowitz, J., Whitehorn, D., Harvey, P.D., DeSmedt, G., Kopala, L.C., 2004. The relationship of neuropsychological test performance with the PANSS in antipsychotic naive, first-episode psychosis patients. Schizophr. Res. 68 (1), 11–19.
Gould, F., McGuire, L.S., Durand, D., Sabbag, S., Larrauri, C., Patterson, T.L., Harvey, P.D., 2015. Self-assessment in schizophrenia: accuracy of evaluation of cognition and everyday functioning. Neuropsychology 29 (5), 675–682.
Green, M.F., Nuechterlein, K.H., Kern, R.S., Baade, L.E., Fenton, W.S., Gold, J.M., Keefe, R.S.E., Mesholam-Gately, R., Seidman, L.J., Stover, E., Marder, S.R., 2008. Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS psychometric and standardization study. Am. J. Psychiatry 165 (2), 221–228.
Green, M.F., Schooler, N.R., Kern, R.S., Frese, F.J., Granberry, W., Harvey, P.D., Karson, C.N., Peters, N., Stewart, M., Seidman, L.J., Sonnenberg, J., Stone, W.S., Walling, D., Stover, E., Marder, S.R., 2011. Evaluation of functionally meaningful measures for clinical trials of cognition enhancement in schizophrenia. Am. J. Psychiatry 168 (4), 400–407.
Harvey, P.D., Serper, M.R., White, L., Parrella, M.J., McGurk, S.R., Moriarty, P.J., Bowie, C., Vadhan, N., Friedman, J., Davis, K.L., 2001. The convergence of neuropsychological testing and clinical ratings of cognitive impairment in patients with schizophrenia. Compr. Psychiatry 42 (4), 306–313.
Harvey, P.D., Velligan, D.I., Bellack, A.S., 2007. Performance-based measures of functional skills: usefulness in clinical treatment studies. Schizophr. Bull. 33 (5), 1138–1148.
Harvey, P.D., Ogasa, M., Cucchiaro, J., Loebel, A., Keefe, R.S.E., 2011. Performance and interview-based assessments of cognitive change in a randomized, double-blind comparison of lurasidone vs. ziprasidone. Schizophr. Res. 127 (1–3), 188–194.
Harvey, P.D., Jacobson, W., Zhong, W., Krarup, G., Melau, M., Forchhammer, H.B., Nordentoft, M., 2017. Determination of a clinically important difference and definition of a responder threshold for the UCSD performance-based skills assessment (UPSA) in patients with major depressive disorder. J. Affect. Disord. 213 (Apr), 105–111.
Heinrichs, R.W., Statucka, M., Goldberg, J., Dermid, Vaz.S., 2006. The University of California performance skills assessment (UPSA) in schizophrenia. Schizophr. Res. 88 (1–3), 135–141.
Higuchi, Y., Sumiyoshi, T., Seo, T., Suga, M., Takahashi, T., Nishiyama, S., Komori, Y., Kasai, K., Suzuki, M., 2017. Associations between daily living skills, cognition, and real-world functioning across stages of schizophrenia; a study with the schizophrenia cognition rating scale Japanese version. hizophr. Res. Cogn. 7 (Feb) Sc, 13–18.
Hinkin, T.R., 1995. A review of scale development in the study of behavior in organizations. J. Manag. 21 (5), 967–988 1995.
Hofer, A., Niedermayer, B., Kemmler, G., Rettenbacher, M.A., Trebo, E., Widschwendter, C.G., Fleischhacker, W.W., 2007. Cognitive impairment in schizophrenia: clinical ratings are not a suitable alternative to neuropsychological testing. Schizophr. Res. 92 (1–3), 126–131.
Javitt, D.C., Buchanan, R.W., Keefe, R.S., Kern, R., McMahon, R.P., Green, M.F., Lieberman, J., Goff, D.C., Csernansky, J.G., McEvoy, J.P., Jarskog, F., Seidman, L.J., Gold, J.M., Kimhy, D., Nolan, K.S., Barch, D.S., Ball, M.P., Robinson, J., Marder, S.R., 2012. Effect of the neuroprotective peptide davunetide (AL-108) on cognition and functional capacity in schizophrenia. Schizophr. Res. 136 (1–3), 25–31.
Kane, J.M., D'Souza, D.C., Patkar, A.A., Youakim, J.M., Tiller, J.M., Yang, R., Keefe, R.S., 2010. Armodafinil as adjunctive therapy in adults with cognitive deficits associated with schizophrenia: a 4-week, double-blind, placebo-controlled study. J. Clin. Psychiatry. 71 (11), 1475–1481.
Keefe, R.S.E., Goldberg, T.E., Harvey, P.D., Gold, J.M., Poe, M., Coughenour, L., 2004. The brief assessment of cognition in schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. Schizophr. Res. 68 (2–3), 283–297.
Keefe, R.S., Poe, M., Walker, T.M., Kang, J.W., Harvey, P.D., 2006. The schizophrenia cognition rating scale: an interview-based assessment and its relationship to cognition, real-world functioning, and functional capacity. Am. J. Psychiatry 163 (3), 426–432.
Keefe, R.S., Meltzer, H.A., Dgetluck, N., Gawryl, M., Koenig, G., Moebius, H.J., Lombardo, I., Hilt, D.C., 2015a. Randomized, double-blind, placebo-controlled study of Encenicline, an α7 nicotinic acetylcholine receptor agonist, as a treatment for cognitive impairment in schizophrenia. Neuropsychopharmacology 40 (13), 3053–3060.
Keefe, R.S.E., Davis, V.G., Spagnola, N., Hilt, D., Dgetluck, N., Ruse, S., Patterson, T.D., Narasimhan, M., Harvey, P.D., 2015b. Reliability, validity and treatment sensitivity of the schizophrenia cognition rating scale. Eur. Neuropsychopharmacol. 25 (2), 176–184.
Keefe, R.S.E., Davis, V.G., Atkins, A.S., Vaughan, A., Patterson, T., Narasimhan, M., Harvey, P.D., 2016. Validation of a computerized test of functional capacity. Schizophr. Res. 175 (1–3), 90–96.
Keefe, R.S.E., Davis, V.G., Harvey, P.D., Atkins, A.S., Haig, G.M., Hagino, O., Marder, S., Hilt, D.C., Umbricht, D., 2017. Placebo response and practice effects in schizophrenia cognition trials. JAMA Psychiatry 74 (8), 807–814.
Kemp, A.S., Schooler, N.R., Kalali, A.H., Alphs, L., Anand, R., Awad, G., Davidson, M., Dubé, S., Ereshefsky, L., Gharabawi, G., Leon, A.C., Lepine, J.P., Potkin, S.G., Vermeulen, A., 2010. What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? Schizophr. Bull. 36 (3), 504–509.
Krishnan, K.K.R., Levy, R.M., Wagner, H.R., Chen, G., Gersing, K., Doraiswamy, P.M., 2001. Informant-rated cognitive symptoms in normal aging, mild cognitive impairment, and dementia: initial development of an informant-rated screen (brief cognitive scale) for mild cognitive impairment and dementia. Psychopharmacol. Bull. 35 (3), 79–88.
Loewenstein, D.A., Argüelles, S., Bravo, M., Freeman, R.Q., Argüelles, T., Acevedo, A., Eisdorfer, C., 2001. Caregivers' judgments of the functional abilities of the Alzheimer's disease patient: a comparison of proxy reports and objective measures. J. Gerontol. B. Psychol. Sci. Soc. Sci. 56 (2), P78–P84.
Markova, I.S., Berrios, G.E., 1992. The assessment of insight in clinical psychiatry: a new scale. Acta Psychiatr. Scand. 86 (2), 159–164.
Mausbach, B.T., Harvey, P.D., Goldman, S.R., Jeste, D.V., Patterson, T.L., 2007. Development of a brief scale of everyday functioning in persons with serious mental illness. Schizophr. Bull. 33 (6), 1364–1372.
Mazhari, S., Ghafaree-Nejad, A.R., Soleymani-Zade, S., Keefe, R., 2017. Validation of the Persian version of the schizophrenia cognition rating scale (SCoRS) in patients with schizophrenia. Asian J. Psychiatr. 27, 12–15.
Menditto, A.A., Wallace, C.J., Liberman, R.P., VanderWal, J., Jones, N.T., Stuve, P., 1999. Functional assessment of independent living skills. Psychiatr. Rehabilitation Skills. 3 (2), 200–219.
Nielsen, R.E., Lindström, E., Telléus, G.K., Levander, S., 2014. Is the PANSS cognitive scale measuring cognition? Nord. J. Psychiatry. 68 (8), 573–578.
Nunnally, J.C., 1978. Psychometric Theory. second ed. McGraw-Hill, New York.
Ouzir, M., Azorin, J.M., Adida, M., Boussaoud, D., Battas, O., 2012. Insight in schizophrenia: from conceptualization to neuroscience. Psychiatry Clin. Neurosci. 66 (3), 167–179.
Patterson, T.L., Goldman, S., McKibbin, C.L., Hughs, T., Jeste, D.V., 2001. UCSD performance-based skills assessment: development of a new measure of everyday functioning for severely mentally ill adults. Schizophr. Bull. 27 (2), 235–245.
Price, J., Mueller, C., 1986. Handbook of Organizational Measurement. Pitman, Massachusetts.
Rodriguez-Jimenez, R., Bagney, A., Mezquita, L., Martinez-Gras, I., Sanchez-Morla, E.M., Mesa, N., Ibañez, M.I., Diez-Martin, J., Jimenez-Arriero, M.A., Lobo, A., Santos, J.L., Palomo, T. PARG, 2013. Cognition and the five-factor model of the positive and negative syndrome scale in schizophrenia. Schizophr. Res. 143 (1), 77–83.
Sabbag, S., Twamley, E.M., Vella, L., Heaton, R.K., Patterson, T.L., Harvey, P.D., 2011. Assessing everyday functioning in schizophrenia: not all informants seem equally informative. Schizophr. Res. 131 (1–3), 250–255.
Schneider, L.C., Struening, E.L., 1983. SLOF: a behavioral rating scale for assessing the mentally ill. Soc. Work Res. Abstr. 19 (3), 9–21.
Shimada, T., Ohori, M., Inagaki, Y., Shimooka, Y., Sugimura, N., Ishihara, I., Yoshida, T., Kobayashi, M., 2018. A multicenter, randomized controlled trial of individualized occupational therapy for patients with schizophrenia in Japan. PLoS One 13 (4), e0193869.
Silberstein, J., Harvey, P.D., 2019. Impaired introspective accuracy in schizophrenia: an independent predictor of functional outcomes. Cogn. Neuropsychiatry. 24 (1), 28–39.
Silberstein, J.M., Pinkham, A.E., Penn, D.L., Harvey, P.D., 2018. Self-assessment of social cognitive ability in schizophrenia: association with social cognitive test performance, informant assessments of social cognitive ability, and everyday outcomes. Schizophr. Res. 199, 75–82 Sept.
Strassnig, M., Rosenfeld, A., Harvey, P.D., 2018. Tardive dyskinesia: motor system impairments, cognition and everyday functioning. CNS Spectr 23 (6), 370–377.
Takahashia, T., Higuchia, Y., Komoria, Y., Nishiyamaa, S., Nakamuraa, M., Sasabayashia, D., Nishikawaa, Y., Sumiyoshi, T., Suzukia, M., 2017. Quality of life in individuals with attenuated psychotic symptoms: possible role of anxiety, depressive symptoms, and socio-cognitive impairments. Psychiatry Res. 257, 431–437.
U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), 2017. Multiple endpoints in clinical trials: Guidance for Industry https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm536750.pdf.
Vesterager, L., Christensen, T.Ø., Olsen, B.B., Krarup, G., Melau, M., Forchhammer, H.B., Nordentoft, M., 2012. Cognitive and clinical predictors of functional capacity in patients with first episode schizophrenia. Schizophr. Res. 141 (2–3), 251–256.
Vita, A., Deste, G., Barlati, S., De Peri, L., Giambra, A., Poli, R., Keefe, R.S., Sacchetti, E., 2013. Interview-based assessment of cognition in schizophrenia: applicability of the schizophrenia cognition rating scale (SCoRS) in different phases of illness and settings of care. Schizopr. Res. 146 (1–3), 217–223.