Cognition Assessment in Virtual Reality

Validity and feasibility of a novel virtual reality test for real-life cognitive functions in mood disorders and psychosis spectrum disorders

Miskowiak, Kamilla W.; Jespersen, Andreas E.; Kessing, Lars, V; Aggestrup, Anne Sofie; Glenthoj, Louise B.; Nordentoft, Merete; Ott, Caroline, V; Lumbye, Anders

Published in:
Journal of Psychiatric Research

DOI:

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Cognition Assessment in Virtual Reality: Validity and feasibility of a novel virtual reality test for real-life cognitive functions in mood disorders and psychosis spectrum disorders

Kamilla W. Miskowiak, Andreas E. Jespersen, Lars V. Kessing, Anne Sofie Aggestrup, Louise B. Glenthøj, Merete Nordentoft, Caroline V. Ott, Anders Lumbye

A R T I C L E   I N F O

Keywords: Virtual reality, Cognition Assessment, Bipolar disorder, Psychosis

A B S T R A C T

There is a pressing need for measures of real-life cognitive functioning in patients with mood or psychotic disorders in clinical settings and treatment trials targeting cognition. We developed the first immersive virtual reality cognitive assessment tool, the Cognition Assessment in Virtual Reality (CAVIR), which assesses verbal memory, processing speed, attention, working memory and planning skills in an interactive virtual reality kitchen scenario. This study investigates the sensitivity and validity of the CAVIR for cognitive impairments in mood and psychotic disorders and its association with functioning and neuropsychological performance. Symptomatically stable patients with mood disorders (MD; n = 40) or psychosis spectrum disorders (PSD; n = 41) and healthy control participants (HC; n = 40) completed the CAVIR and standard neuropsychological tests and were rated for clinical symptoms and daily functioning. We found that the CAVIR was sensitive to cognitive impairments across MD and PSD with large effect sizes (MD: F(73) = 11.61, p < .01, ηp² = 0.14; PSD: F(72) = 18.24, p < .001, ηp² = 0.19). There was a moderate to strong positive correlation between performance on the CAVIR and on neuropsychological tests (r(121) = 0.58, p < .001), which prevailed after adjustment for age, years of education and verbal IQ (B = 0.67, p < .001). Lower CAVIR scores correlated moderately with more observer-rated and performance-based functional disability (r(121) = 0.30, p < .01 and r(68) = 0.44, p < .001, respectively), also after adjustment for age, years of education and verbal IQ (B = 0.03, p < .001). In conclusion, the CAVIR is a sensitive and valid instrument for measuring real-life cognitive impairments in mood and psychotic disorders. After further psychometric assessments, the CAVIR can be implemented in clinical settings and trials targeting cognition.

1. Introduction

Cognitive impairment is a core feature of several neuropsychiatric disorders, including mood disorders (MD; unipolar and bipolar disorder) and psychosis spectrum disorders (PSD) (Bora et al., 2013; Bourne et al., 2013; Rund, 1998). The impairment is evident across several cognitive domains and persists during asymptomatic illness phases. Cognitive impairments lead to poorer prognosis and functional disability (Tse et al., 2014) of which reduced workforce capacity comprises the largest socio-economic burden of MD (Olesen et al., 2012) and PSD (Chong et al., 2016). Cognitive impairment is therefore an urgent treatment target to improve outcome for patients and reduce societal costs (Bowie and Harvey, 2006; Miskowiak et al., 2017). Nevertheless, there are no available treatments with robust and replicated efficacy on cognition in MD (Miskowiak et al., 2017) and no biological treatments with cognitive benefits in PSD (Goff et al., 2011). While cognitive remediation produces moderate benefits in PSD overall, a large proportion of patients show no improvement (Wykes et al., 2011).

The limited success of cognition trials is partly due to broad methodological problems (Miskowiak et al., 2017). A key challenge is that

* Corresponding author. Department of Psychology, University of Copenhagen, Copenhagen, Denmark.
E-mail address: kamilla.miskowiak@regionh.dk (K.W. Miskowiak).

Received 11 July 2021; Received in revised form 8 November 2021; Accepted 10 December 2021
Available online 12 December 2021
0022-3956/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
neuropsychological outcomes—the primary outcome of cognition trials—have limited ecological validity (Miskowiak et al., 2017). Neuropsychological tests are administered in a controlled environment with specific instructions, which retains little resemblance to real-life cognitive challenges. Accordingly, performance on neuropsychological tests accounts for only 5–21% variance in patients’ daily functioning (Van der Elst et al., 2008). Hence, the use of neuropsychological tests as primary outcomes provides limited insight into whether candidate treatments aid real-life functioning (Lewandowski et al., 2017; Miskowiak et al., 2017; Ott et al., 2016; Ruse et al., 2014b); an important prerequisite for their approval. The International Society for Bipolar Disorder (ISBD) Targeting Cognition Task Force therefore recommended that functional capacity is included as secondary outcome in cognition trials (Miskowiak et al., 2017). Nevertheless, the few trials that found treatment-related effects on patients’ neuropsychological functions showed no corresponding improvements in the secondary functional outcomes (Lewandowski et al., 2017; Miskowiak et al., 2014; Ott et al., 2016). An alternative explanation to the poor ecological validity of neuropsychological tests is that available measures of functioning also have limitations. Observer-rated measures, such as the Functional Assessment Short Test (FAST) (Rosa et al., 2007), are more closely related to subjective cognitive difficulties and mood symptoms than to objective cognitive performance (Ott et al., 2019). In contrast, performance-based measures, such as the Brief University of California, San Diego Performance-based Skills Assessment (UPS A) (Mausbach et al., 2010), correlate more consistently with objective neuropsychological performance but have ceiling effects in MD and PSD and retain transcultural problems (Miskowiak et al., 2017; Ott et al., 2019; Østergaard Christensen et al., 2014). A more general limitation is that these measures do not directly assess patients’ real-world functioning. These limitations of neuropsychological and functional measures highlight a need for novel integrative outcome measures that capture patients’ cognitive functioning in real-life scenarios.

Virtual reality (VR) platforms enable simulation of engaging, naturalistic cognitive challenges while maintaining a controlled experimental environment. Virtual reality is therefore a promising assessment modality for ecologically valid measurement of real-life cognitive functioning. In keeping with this, the ISBD Targeting Cognition Task Force highlighted the Virtual Reality Functional Capacity Assessment Tool (VRF-CAT) as a promising tool to assess real-time functional capacity (Miskowiak et al., 2017; Ruse et al., 2014a). However, we have no equivalent VR tools to measure patients’ real-life cognitive functions, including verbal memory, attention, and planning abilities. To address this need, we developed a self-administered immersive VR test, the Cognition Assessment in Virtual Reality (CAVIR) that measures several cognitive domains within a virtual kitchen scenario. This study investigated: (I) the sensitivity of the CAVIR to real-life cognitive impairments in a sample of symptomatically stable patients with MD or PSD, (II) the concurrent and ecological validity of the CAVIR, and (III) the feasibility of the CAVIR in these patient groups.

2. Materials and methods

2.1. Participants and recruitment

Participants with MD or PSD and HC were recruited between May 2019–February 2020. Patients with MD were recruited from outpatient clinics at the Psychiatric Centre Copenhagen and our longitudinal Bipolar Illness Onset (BIO) study (Reissing et al., 2017). Patients with PSD were recruited from early intervention outpatient clinics in the Capital region of Denmark. Healthy controls (HC) were recruited through the BIO study and the Blood Bank at Copenhagen University Hospital.

Eligible participants were 18–60 years of age. Patients with MD had an ICD-10 diagnosis of either unipolar disorder (UD, n = 12) or bipolar disorder (BD, n = 28) and were in full or partial remission, as reflected by scores ≤14 on the Hamilton Depression Rating Scale 17 items (HRDS-17) and the Young Mania Rating Scale (YMRS). Patients with PSD had an ICD-10 diagnosis of either schizophrenia (n = 15), schizotypal disorder (n = 24) or non-affective psychosis (n = 2). Healthy controls were free of any personal or first-degree family history of psychiatric illness. Diagnostic interviews were conducted with the Present State Examination, Schedules for Clinical Assessment in Neuropsychiatry or Mini International Neuropsychiatric Interview.

Exclusion criteria were current substance use disorder, neurological disorder, severe somatic illness and dyslexia. Patients were excluded if they had a daily use of benzodiazepines corresponding to >22.5 mg oxazepam or had electroconvulsive therapy (ECT) within the past three months. The investigation was carried out in accordance with the latest version of the Declaration of Helsinki. The study was approved by the data protection agency in the Capital Region of Denmark (VD-2018-468). The study design was reviewed by an appropriate ethical committee which stated that the study required no approval because it involved no invasive procedures. Written informed consent of the participants was obtained after the nature of the procedures had been fully explained.

2.2. Procedure

Participants attended the Copenhagen University Hospital, Rigshospitalet, or their early intervention outpatient clinic. Assessments involved completion of the CAVIR, a neuropsychological test battery, and functional measures. Patients with MD and HC also underwent mood ratings with the HDRS-17 and YMRS. For patients with PD, positive symptoms were assessed using the Assessment of Positive Symptoms (SAPS) and negative symptoms using the Scale for the Assessment of Negative Symptoms (SANS).

2.3. Materials

2.3.1. Cognition Assessment in Virtual Reality

The CAVIR is an immersive VR test of daily-life cognitive functions in an interactive VR kitchen scenario administered on a standalone head-mounted Oculus Go 32 GB portable headset running on a 5.5 inch LCD display with a resolution 1280×1440 pixels per eye at 72hz refresh rate. The headset is using a Qualcomm Snapdragon 821 chipset as the main CPU and GPU processing unit. To navigate the environment the participant uses a hand-held controller (https://www.oculus.com/go/) (Fig. 1; Table 1). The CAVIR is self-administered with a 15 min duration and was developed by AL in collaboration with KWM and CVO. Before putting on the VR headset, participants were instructed that CAVIR involves five tasks in a simulated kitchen and that they would receive instructions in the headset before the tasks begin and they were shown how to use the controller (for details see the supplementary materials). Through headphones, participants were instructed by a pre-recorded voice to carry out five tasks involved in planning and preparing a meal. These tasks measure verbal memory, executive functions, processing speed, working memory and attention, respectively. For further details, see the supplementary materials.

Verbal learning is assessed in task 1, in which participants are shown a list with ingredients and instructed to memorize these and take them out of the fridge (Fig. 1A). Performance is measured by the number of correct ingredients remembered (score range: 0–10). Executive functions are assessed in task 2, in which participants are required to plan and select the order in which to perform different sub-tasks involved in cooking a meal to finish before their guests’ arrival (Fig. 1B). Here, performance is measured the number of correctly placed tasks that ensure timely completion of the tasks (score range: 0–7). Processing speed is assessed in task 3, in which participants place as many correct ingredients as possible in a pot within 90 s based on a key of symbols matching every ingredient (no score range; Fig. 1C). Working memory is assessed in task 4, in which participants observe and memorize the location of cutlery and flatware in the kitchen cupboards and drawers.

K.W. Miskowiak et al.
Performance is measured by the number of drawers opened until all cutlery and flatware is found (lower indicating better performance). Finally, sustained attention is measured in task 5, in which participants are required to repeatedly check the lasagne in the oven in response to a specific combination of visual and auditive cues while ignoring irrelevant stimuli (Fig. 1E). Performance is measured by the number of correct responses (score range: 0–20).

### 2.3.2. Standard neuropsychological measures

Neuropsychological function was investigated with the One Touch Stocking of Cambridge (OTS), Spatial Working Memory (SWM) test and Rapid Visual Information Processing (RVP) from CANTAB (Cambridge Cognition Ltd.), Rey Auditory Verbal Learning Test (RAVLT), WAIS-III letter-number sequencing, RBANS Coding and digit span, verbal fluency (‘d’ and ‘s’) and Trail Making A and B. Verbal intelligence was estimated with the Danish Adult Reading Test (DART).

### 2.3.3. Functional assessments

Functioning was assessed with the FAST, an observer-based rating scale and the UPSA-B, a performance-based measure of functioning, assessing Financial and Communication skills. Both tools are recommended by the ISBD Targeting Cognition Task force (Miskowiak et al., 2017).

### 2.3.4. Assessment of the experience in the virtual reality environment

To assess user-experience and the degree to which participants experienced ‘presence’ in the CAVIR environment, participants filled out a shortened version of the Presence Questionnaire (PQ). Participants also filled out the VR Simulation Sickness Questionnaire (VRSSQ). The PQ and VRSSQ were translated from English to Danish and back translated. This work was completed half-way through the project and the questionnaires were therefore only given about half of the participants.

### 2.3.5. Statistical analyses

Scores on the CAVIR tests and neuropsychological tests were z-transformed based on the mean and standard deviation of HC. Outlying z-scores (>4 SDs below HC mean) were truncated to z = -4.0 to limit the impact of extreme scores while still allowing variability in the data in line with previous studies from our group (Jensen et al., 2016). For tests, in which lower scores indicated better performance (i.e., latency in the CAVIR task 2; TMT tests, RVP, SWM and OTS), the scores were inverted.
before standardization to ensure that all scales had the same direction. For the CAVIR, five cognitive domains were calculated by averaging the z-transformed scores within each of the five sub-tasks (see Table 1 for an overview of the CAVIR cognitive domains). A global CAVIR composite score was then calculated by averaging the five domains. Data was excluded for participants who had misunderstood individual CAVIR tasks, as assessed by a performance >3 SD below mean of their respective group and verbal reports. For participants with missing CAVIR data, the global CAVIR composite was calculated by averaging the remaining domains. Five cognitive domains were also calculated based on the z-transformed neuropsychological test scores within the respective domains (see Table 1). A global neuropsychological composite score was calculated by averaging these five cognitive domains.

Groups were compared for demographic and clinical variables using independent sample t-tests and for gender distribution using X² tests. Comparisons were conducted separately for MD vs. HC and for PSD vs. HC, because the main hypotheses involved separate comparisons for these groups. Sensitivity of the CAVIR to cognitive impairment was assessed with one-way analysis of variance (ANOVA) for MD and PSD compared with HC, respectively, with the CAVIR global or domain scores as dependent variables and group as independent variable. Significant group differences were followed up with ANCOVA with adjustment for any demographic or clinical variables, on which patients and HC differed. Similar ANOVAs and ANCOVAs were conducted to investigate group differences in neuropsychological and functional measures. Concurrent and ecological validity of the CAVIR was assessed with Pearson’s correlations across the entire sample between (I) the CAVIR and neuropsychological global and domain scores and (II) the CAVIR global composite and measures of functioning, respectively. Statistical analyses were conducted with the Statistical Package for Social Sciences (version 25; IBM Corporation, Armonk, New York). Statistical significance was set to an alpha-level of p < .05 (two tailed).

### Table 2
Demographic and clinical variables, functional capacity and composite scores for the Cognition Assessment in Virtual Reality (CAVIR) tests and corresponding cognitive domains for patients with mood disorders (MD; n = 40) or psychosis spectrum disorders (PSD; n = 41) and healthy controls (HC; n = 40).

<table>
<thead>
<tr>
<th>Group</th>
<th>Group comparisons</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HC (N = 40)</td>
<td>MD (UD/BD) (N = 40)</td>
</tr>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Demographic and clinical variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>30.4 (9.6)</td>
<td>15.7 (1.8)</td>
</tr>
<tr>
<td>Years of education</td>
<td>35.3 (11.7)</td>
<td>14.9 (2.3)</td>
</tr>
<tr>
<td>Sex, n (% female)</td>
<td>22.4 (3.9)</td>
<td>13.7 (2.3)</td>
</tr>
<tr>
<td>HDRS-17 baseline</td>
<td>.40**</td>
<td>.028*</td>
</tr>
<tr>
<td>YMRS baseline</td>
<td>.001**</td>
<td>.067</td>
</tr>
<tr>
<td>Est. premorbid ability</td>
<td>.906</td>
<td>.001**</td>
</tr>
<tr>
<td>SANS</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SANS psychotic A: G</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SANS disorganised</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>CAVIR domains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention²</td>
<td>0.0 (1.0)</td>
<td>–0.4 (1.1)</td>
</tr>
<tr>
<td>Processing speed⁴</td>
<td>0.0 (1.0)</td>
<td>–1.2 (1.5)</td>
</tr>
<tr>
<td>Verbal learning/memory</td>
<td>0.0 (0.7)</td>
<td>–0.7 (1.2)</td>
</tr>
<tr>
<td>Working memory</td>
<td>0.0 (0.9)</td>
<td>–0.6 (1.5)</td>
</tr>
<tr>
<td>Executive functions</td>
<td>0.0 (0.8)</td>
<td>–0.3 (0.7)</td>
</tr>
<tr>
<td>Global cognition</td>
<td>0.0 (0.5)</td>
<td>–0.6 (0.7)</td>
</tr>
<tr>
<td>Cognitive domains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention²</td>
<td>0.0 (0.6)</td>
<td>–0.1 (0.6)</td>
</tr>
<tr>
<td>Processing speed⁴</td>
<td>0.0 (0.9)</td>
<td>–0.8 (1.0)</td>
</tr>
<tr>
<td>Verbal learning/memory</td>
<td>0.0 (0.8)</td>
<td>–0.5 (1.1)</td>
</tr>
<tr>
<td>Working memory⁰</td>
<td>0.0 (0.8)</td>
<td>–0.5 (0.8)</td>
</tr>
<tr>
<td>Executive functions</td>
<td>0.0 (0.7)</td>
<td>–0.4 (0.9)</td>
</tr>
<tr>
<td>Global cognition</td>
<td>0.0 (0.5)</td>
<td>–0.5 (0.6)</td>
</tr>
<tr>
<td>Functioning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAST</td>
<td>1.8 (2.8)</td>
<td>15.2 (9.3)</td>
</tr>
<tr>
<td>UPSA-B</td>
<td>–</td>
<td>81.8 (9.3)</td>
</tr>
<tr>
<td>Virtual reality questionnaires</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PQ (‘presence’)</td>
<td>5.2 (0.6)</td>
<td>4.9 (0.7)</td>
</tr>
<tr>
<td>VRSSQ (simulation sickness)</td>
<td>0.2 (0.3)</td>
<td>0.3 (0.2)</td>
</tr>
</tbody>
</table>


² Missing data for n = 1 MD patient.
⁴ Missing data for the CAVIR attention task (task 5) n = 1 PSD patient.
⁵ Missing data for the CAVIR processing speed task (task 3) n = 1 MD patient and n = 4 PSD patients.
⁶ Missing data for the Rapid Visual Processing task n = 12 MD patients and n = 4 PSD patients.
⁰ Missing data for the Spatial Working Memory Task n = 13 MD patients and n = 1 PSD patient.
⁷ Missing data for the One-Touch Stockings of Cambridge n = 29 MD patients and n = 13 HCs.
⁸ Missing data for n = 16 patients with MD and all HCs.
3. Results

3.1. Demographics and clinical variables

Clinical and demographic data is presented in Table 2. We included 40 participants with MD, 41 participants with PSD and 40 HC. The MD and HC groups were comparable for gender, verbal IQ and education levels (p-values > .07). The MD group was older (t = −2.03, df = 78, p = .046) and displayed more subsyndromal depression (t = 6.81, df = 49.89, p < .001) and mania symptoms (t = −2.63, df = 59.81, p = .01). Compared to HC, the PSD group was younger (t = 3.84, df = 79, p < .001), included more males (X² (1, 81) = 4.50, p = .03, Φ = 0.24), had lower verbal IQ (t = 2.15, df = 79 p = .03) and fewer years of education (t = 4.34, df = 79 p < .001).

3.2. Sensitivity of the CAVIR to real-life cognitive impairments

Table 2 and Fig. 2 display CAVIR performance of patients and HC. Patients with MD showed lower CAVIR global cognition scores with a large effect size (F(1,78) = 21.3, p < .001, η² = 0.21; Table 2, Fig. 2), which prevailed after adjustment for age and mood symptoms (F(1,78) = 11.90, p = .001, η² = 0.14). For CAVIR subtests, patients with MD displayed impaired verbal memory (F(1,78) = 9.46, p = .003, η² = 0.11), processing speed (F(1,78) = 17.61, p < .001, η² = 0.19) and working memory (F(1,78) = 4.51, p = .04, η² = 0.06), which prevailed for processing speed and verbal learning after adjustment for age and mood symptoms (p-values < .02). Patients showed a trend-level impairment on the CAVIR executive function subtest (F(1,78) = 3.92, p = .051, η² = 0.05), which rendered significant after adjustment for age and mood symptoms (F(1,78) = 6.18, p = .02, η² = 0.08). Patients with MD showed no impairment on the CAVIR attention subtest (p = .84).

Patients with PSD showed large impairment on the CAVIR global composite (F(1,78) = 15.56, p < .001, η² = 0.17), which prevailed after adjustment for age, gender, education and verbal IQ (F(1,79) = 15.47, p < .001, η² = 0.17). Their performance was also impaired on CAVIR verbal memory (F(1,78) = 7.50, p = .008, η² = 0.09), processing speed (F(1,78) = 7.15, p = .009, η² = 0.09) and executive functions subtests (F(1,78) = 9.78, p = .002, η² = 0.11), which prevailed after controlling for age, gender, education and verbal IQ (p-values < .009). There was a non-significant trend towards working memory impairment in PSD (F(1,78) = 3.06, p = .08, η² = 0.04), and no group difference in sustained attention (F(1,78) = 0.91, p = .34, η² = 0.01), also after controlling for demographic variables (p-values ≥ .15). Posthoc analysis showed that 67% of patients achieved a maximum score on the attention subtest.

3.3. Group differences on neuropsychological and functional measures

Patients with MD showed impaired global neuropsychological performance with large effect size (F(1,78) = 12.42, p = .001, η² = 0.14; Table 2). Across individual domains, they exhibited impaired verbal learning and memory (F(1,78) = 5.37, p = .02, η² = 0.06), processing speed (F(1,78) = 14.63, p = .001, η² = 0.16), working memory (F(1,78) = 8.69, p = .004, η² = 0.100) and executive functions (F(1,78) = 6.78, p = .011, η² = 0.08), which rendered non-significant after adjustment for age and mood symptoms (ps > .06). They showed no impairment within sustained attention (F(1,78) = 0.29, p = .59, η² = 0.004; Table 2). Patients with PSD impaired global neuropsychological scores than HC (F(1,79) = 33.93, p < .001, η² = 0.30) and impairments within verbal learning and memory (F(1,79) = 9.13, p = .003, η² = 0.10), processing speed (F(1,79) = 41.71, p < .001, η² = 0.35), working memory (F(1,79) = 6.15, p = .015, η² = 0.07) and executive functions (F(1,79) = 32.09, p < .001, η² = 0.29) that prevailed after controlling for age, gender, education and premorbid verbal IQ (p-values ≤ .01). Patients showed only a non-significant trend towards impaired sustained attention (F(1,78) = 3.66, p = .059, η² = 0.04).

The MD group exhibited considerable everyday life problems on the FAST (F(1,78) = 74.68, p < .001, η² = 0.49; Table 2), which prevailed after adjustment for age and mood symptoms (p = .019). Patients with PSD also presented with functional disability on the FAST (F(1,78) = 125.41, p < .001, η² = 0.61), which were larger than the functional disability in MD patients (F(1,78) = 5.72, p = .019, η² = 0.07) and prevailed after adjustment for age, gender, education and verbal IQ (p-values ≤ .02). The UPSA-B was not completed by the HC from the BIO study, which impeded comparisons between patients and HC. However for comparison, patients with MD or PSD scored 1.8 SD and 2 SD lower than middle-aged HC in a previous study (Depp et al., 2009). Notably, UPSA-B scores were comparable between MD and PSD patients (F(1,63) = 0.08, p = .79, η² = 0.001).

3.3.1. Concurrent and ecological validity of the CAVIR test

Performance on the CAVIR correlated strongly with neuropsychological performance across the entire sample (r(121) = 0.58, p < .001; see Table 3), which prevailed after adjustment for age, education and verbal IQ (B = 0.67, p < .001). Scores on CAVIR subtests also correlated with corresponding neuropsychological domain scores for verbal memory (r(121) = 0.29, p < .001), processing speed (r(116) = 0.51, p < .001), working memory (r(121) = 0.31, p = .001), and executive functions (r(121) = 0.26, p = .004), which remained significant after adjustment for age, education and verbal IQ (ps < .010). There was a non-significant trend towards a correlation between the CAVIR and...
neuropsychological sustained attention scores (p = .07).

Lower global CAVIR scores were associated with more functional disability, as reflected by moderate correlations with higher FAST scores (r(121) = −0.30, p < .001) and lower UPSA-B scores (r(68) = 0.44, p < .001), which prevailed after adjustment for age, education and verbal IQ (B = −0.03, p < .001 and B = 0.03, p < .001, respectively). Notably, lower global scores on traditional neuropsychological tests were also associated with higher FAST scores (r(121) = −0.45, p < .001) and lower UPSA-B scores (r(68) = 0.52, p < .001).

### 3.3.2. Usability and tolerability of the CAVIR test

Fifty-two (46%) of the 121 participants were asked to complete the PQ and VRSSQ (HC: n = 19; AD: n = 19; PD: n = 14). Participants reported a high presence of the CAVIR environment (M = 5.0, SD = 0.8 of the maximum score of 7.0; see supplementary table) and good usability of headset and controller (M = 5.7, SD = 0.9 of the maximum score of 7.0). The CAVIR produced only minimal ‘simulation sickness’, indicated by low scores on the VRSSQ across all groups (M = 0.3, SD = 0.4; Table 2 and supplementary table).

### 4. Discussion

This study investigated the validity and feasibility of the first available comprehensive virtual reality cognition test, the CAVIR, in a sample of symptomatically stable patients with mood or psychosis spectrum disorders and healthy controls. The CAVIR was sensitive to cognitive impairments across mood and psychosis spectrum disorders with moderate to large effect sizes. The instrument showed high concurrent validity, as indicated by strong correlations between CAVIR and on neuropsychological test scores, which prevailed after adjustment for age, years of education and IQ. The ecological validity of the CAVIR was adequate, as indicated by moderate correlations with observer-rated and performance-based functional disability, also after adjustment for age, years of education and verbal IQ. Participants rated the CAVIR as highly engaging, relevant for their daily lives, and as producing high experience of presence and minimal simulation sickness.

The robust associations between performance on the CAVIR and both neuropsychological tests and measures of daily functioning is consistent with studies of two VR functional capacity and executive function tests. In schizophrenia, performance on the computer-based, non-immersive VRFCAT correlated moderately with scores on the MATRICS Consensus Cognitive Battery and performance-based functional capacity (Keefe et al., 2016; Ruse et al., 2014b; Ventura et al., 2020). Similarly, a VR test for executive function, the Virtual Multiple Errands Test, showed strong correlations with a neuropsychological executive function measure and self-reported ability to carry out activities of daily living in post-stroke patients (Rand et al., 2009). Together, the findings indicate good concurrent validity of these VR tools for the assessment of functional capacity and executive function in neuropsychiatric populations. Nevertheless, no previous study investigated the validity of a comprehensive VR cognition test measuring several cognitive domains, including memory, working memory, planning skills and psychomotor speed. Given its adequate concurrent and ecological validity, the CAVIR meets the existing need for a broadly ecologically valid cognition test battery in cognition intervention trials and clinical settings.

An important distinction regarding the ecological validity is between construct-driven and function-led VR tests (Burgess et al., 2006; Parsons, 2015). Construct-driven VR tests provide insight into the specific cognitive domains but may have little predictive ability regarding patients’ real-world functioning. In contrast, function-led tests capture how well patients tackle everyday life challenges but do not elucidate the specific cognitive domains (Burgess et al., 2006; Parsons, 2015). The CAVIR has both construct-driven and function-led ecological validity. The memory and executive function subtests are primarily function-led, capturing how well patients fare in daily life with memorising ingredients to be taken out of a fridge and with planning subtasks involved in preparing a meal, respectively. In contrast, the working memory, psychomotor speed and sustained attention tests are more construct-driven; they resemble classical neuropsychological tests embedded in a real-world kitchen scenario.

The combination of function-led and construct-driven validity in the CAVIR enables assessments of both how well patients tackle daily life challenges and the specific cognitive domains. In cognition trials, the CAVIR may thus elucidate both (i) whether the candidate treatments improve patients’ real-life cognitive functioning and (ii) the profile of treatment effects across the different cognitive domains. Subject to further psychometric evaluations, this would make the CAVIR an attractive primary or co-primary outcome for cognition intervention trials that would solve the challenge that treatments should demonstrate efficacy on both cognition and functioning (Miskowiak et al., 2017). Nevertheless, research into VR cognition assessment is at a too early stage for any suggestions regarding its potential to substitute neurocognitive tests. The recommendation would therefore be to include both types of measures in cognition trials and to further investigate their associations with one another and measures of real-world functioning. In clinical settings, the CAVIR could help healthcare professionals gain insight into (i) their patients’ cognitive abilities in real-world cognitively challenging situations and (ii) which aspects of patients’ daily cognitive abilities are mostly affected. The CAVIR would represent a major leap forward from the commonly used observer-based or self-report measures of cognition and functioning that are coloured by depressive symptoms and variable insight (Peterson et al., 2019). As such, the CAVIR can provide a solid basis for healthcare professionals to examine the effects of changing medication or unhealthy lifestyle factors and to help patients plan how to compensate and train their cognitive abilities in daily settings, in line with the recommendations by the ISBD Targeting Cognition Task Force (Miskowiak et al., 2018). Importantly, the CAVIR is also cost-effective because it is largely self-administered and involves relatively inexpensive equipment (Oculus Go headsets that cost around $200). Work is now underway to test the validity, test-retest reliability and feasibility of the optimized, translated versions of CAVIR. Another next important step is to develop a parallel version for repeated testing in cognition trials.

A strength was the relatively large sample size (n = 121) and the
inclusion of patients with MD and PSD. However, larger samples would be required to determine the validity of the CAVIR for each separate diagnostic group. Another strength was that patients were symptomatically stable. This enabled insight into patients’ persistent cognitive and functional impairments rather than merely impairments during acute mood or psychosis episodes. A limitation was that PSD patients were in an early stage of illness and may not be representative of all patients with these disorders. Further, medication status was not evaluated, and we therefore could not investigate the potential influence of medications on their CAVIR performance. Another limitation was that researchers conducting the assessments were not blinded to participants’ diagnoses, although this is unlikely to have influenced participants’ performance in the fully automatic CAVIR test. The attention task was associated with ceiling effects in our patient sample and will be revised to include greater difficulty levels in a second version of the CAVIR. Notably, clinical implementation of the CAVIR may be difficult in older age patients with no previous experience with VR or computer technology. Hence, VR cognition assessments of older age patients may indicate that cognitive functions are poorer than what is the case. It is therefore necessary to compare VR assessed cognition in older age patients with an (older) age-matched healthy control group. Finally, while the CAVIR scores correlated with both FAST and UPSA-B measures of functioning, the correlations were not stronger than those between traditional neuropsychological test scores and functioning. This indicates that other factors influence patients’ functioning and that these were not fully captured by CAVIR or traditional cognition tests. These factors likely include, but are not limited to, individual differences in coping styles, personality characteristics, metacognitive assumptions and social support.

In conclusion, there is a pressing need for ecologically valid cognition assessment tools that bridge measures of neuropsychological performance with daily functioning. In this report, we show that a novel comprehensive VR cognition assessment tool, the CAVIR, has high validity, sensitivity and feasibility for assessment of real-life cognitive functions in mood and psychotic disorders. The CAVIR provides an attractive alternative to neuropsychological assessments in intervention trials targeting cognition and in clinical settings, as it enables insight into patients’ real-life cognitive functions across several domains. If the CAVIR is shown to also have adequate test-retest reliability, its implementation will provide a basis for better assessment of cognition treatments that aim to improve functioning and quality of life in patients with mood or psychotic disorders.

Author contributions

All authors met all four ICMJE criteria for authorship. Kamilla Miskowiak was involved in the initial conception and design of the study and the development of the CAVIR, analysing the data and wrote the first draft of the article. Anders Lumbye was involved in the development of the CAVIR with the input from Kamilla Miskowiak and Caroline Ott. Anders Lumbye and Caroline Ott were also involved in the interpretation of the data. Andreas Jespersen, Louise Glenthøj, Merete Nordentoft, Anne Sofie Aggestrup, Lars Vedel Kessing contributed to the acquisition and interpretation of the data. All authors and approved of the final version to be published and agree to be accountable for all aspects of the work.

Previous presentation

None.

Location of work and address for reprints

Neurocognition and Emotion in Affective Disorders (NEAD) Group, CADIC, Psychiatric Centre Copenhagen, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Email: kamilla.miskowiak@regionh.dk.

Declaration of competing interest

Kamilla Miskowiak has received consultancy fees from Lundbeck and Janssen-Cilag in the past three years. Lars Vedel Kessing has during recent three years received consultancy fees from Lundbeck. Andreas Jespersen, Caroline Ott, Louise Glenthøj, Merete Nordentoft, Anne Sofie Aggestrup and Anders Lumbye report no conflicts of interest.

Acknowledgments

TrygFonden is acknowledged for financial support for the study (grant no. 127318). Dr Kamilla W. Miskowiak is supported by a Lundbeck Foundation fellowship (grant no. grant no. R215-2015-4121).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpsychires.2021.12.002.

References

in bipolar disorder by the international society for bipolar disorders targeting cognition task force. Bipolar Disord. 19 (8), 614–626.


