**SUPPLEMENTARY MATERIAL**

**eTable 1:** PRISMA 2020 item checklist………..…..…..…..…..…..…..…..…..…..…..…..…..….…..…..…..…..…..….…..…..……………...…page 2-4

**eTable 2:** PRISMA 2020 item checklist Abstract....…..…..…..…..…..…..…..….…..…..…..….…..…..…..…..…..….…..…..…..…..………......page 5

**eTable 3:** Virtual reality equipment and tasks/experiments carried out on the included studies. .…..…..….…..…...……..…..…..…..……..page 6-8

**eTable 4:** Quality assessment on the included studies.…..…..…..…..…..…..…..…..…..…..…..…...…..…..…..…..….…..…..…..….…….......page 9

**This supplementary material has been provided by the authors to give readers additional information about their work.**

**eTable 1: PRISMA 2020 item checklist**

|  |  |  |  |
| --- | --- | --- | --- |
| **Section and topic** | **Item #** | **Checklist item** | **Location where item is reported** |
| **TITLE** | | |  |
| Title | 1 | Identify the report as a systematic review or a meta-analysis | Title |
| **ABSTRACT** | | |  |
| Structured summary | 2 | See the PRISMA 2020 for Abstracts checklist (table 2) | Abstract |
| **INTRODUCTION** | | |  |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | Introduction |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Introduction |
| **METHODS** | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Methods, eMethods 1 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists, and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Methods |
| Search strategy | 7 | Present the full search strategies for all databases, registers, and websites, including any filters and limits used. | Methods |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and, if applicable, details of automation tools used in the process. | Methods |
| Data collection  process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and, if applicable, details of automation tools used in the process. | Methods |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Methods |
| 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding source). Describe any assumptions made about any missing or unclear information. | Methods |
| Study risk of bias  assessment | 11 | Specify the methods used to assess the risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and, if applicable, details of automation tools used in the process. | Methods, eTable 3 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Methods |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Methods |
| 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling missing summary statistics or data conversions. | Methods |
| 13c | Describe any methods used to tabulate or visually display the results of individual studies and syntheses. | Methods |
| 13d | Describe any methods used to synthesise results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Methods |
| 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, metaregression). | Methods |
| 13f | Describe any sensitivity analyses conducted to assess the robustness of the synthesised results.  Reporting bias assessment. | Methods |
| Reporting bias  assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Methods |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome | Methods |
| **RESULTS** | | | |
| Study selection | 16 | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram (see Fig 1). | Results, figure 1 |
| Study characteristics | 17 | Cite each included study and present its characteristics. | Table 1 |
| Risk of bias within studies | 18 | Present assessments of risk of bias for each included study | Table 1 |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and it’s precision (e.g. confidence/credible interval), ideally using structured tables or plots. | Results, Table 1 |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Results |
| 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | N.a. |
| 20c | Present results of all investigations of possible causes of heterogeneity among study results. | N.a. |
| 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesised results | N.a. |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | Results |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | Results |
| **DISCUSSION** | | | |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | Discussion |
| 23b | Discuss any limitations of the evidence included in the review. | Discussion |
| 23c | Discuss any limitations of the review processes used. | Discussion |
| 23d | Discuss the implications of the results for practice, policy, and future research | Discussion |
| **OTHER INFORMATION** | | | |
| Registration and  protocol | 24a | Provide registration information for the review, including the register name and registration number, or state that the review was not registered. | Methods |
| 24b | Indicate where the review protocol can be accessed or state that a protocol was not prepared. | Methods |
| 24c | Describe and explain any amendments to information provided at registration or in the protocol. | N.a. |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | After discussion |
| Competing interests | 26 | Declare any competing interests of review authors | After discussion |
| Availability of data,  code, and other  materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | After discussion |

**eTable 2: PRISMA 2020 item checklist Abstract**

|  |  |  |  |
| --- | --- | --- | --- |
| **Section and topic** | **Item #** | **Checklist item** | **Location/explanation** |
| **TITLE** | | |  |
| Title | 1 | Identify the report as a systematic review | Abstract |
| **BACKGROUND** | | |  |
| Objectives | 2 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Abstract |
| **METHODS** | | |  |
| Eligibility criteria | 3 | Specify the inclusion and exclusion criteria for the review. | Abstract |
| Information sources | 4 | Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched. | Abstract |
| Risk of bias | 5 | Specify the methods used to assess risk of bias in the included studies. | Abstract |
| Synthesis of results | 6 | Specify the methods used to present and synthesise results. | Abstract |
| **Results** | | |  |
| Included studies | 7 | Give the total number of included studies and participants and summarise relevant characteristics of studies. | Abstract |
| Synthesis of results | 8 | Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured). | Abstract |
| **Discussion** | | |  |
| Limitations of evidence | 9 | Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency, and imprecision). | Abstract |
| Interpretation | 10 | Provide a general interpretation of the results and important implications. | Abstract |
| **Other** | | |  |
| Funding | 11 | Specify the primary source of funding for the review. | Specified in the manuscript because of limited space words) |
| Registration | 12 | Provide the registered name and registration number. | Abstract |

**eTable 3: Virtual reality equipment and tasks/experiments carried out on the included studies.**

|  |  |  |
| --- | --- | --- |
| **First author, year** | **VR equipment** | **Tasks/experiments carried out** |
| (Gould et al. 2007) | VR Town. No details given. | On the first day, participants were oriented in the village for 20 minutes and then had 30 minutes to navigate through it using a computer-generated route. Then they had to find at least three predetermined destinations in two 3-minute practice sessions. If they were succesful, the training session ended; otherwise, they had 30 minutes to explore and familiarize with the locations. Three days after, they could explore the village for 20 minutes to reacquaint themselves with the task. 2-4 hours later, they were assessed on their memory of the village based on the number of locations found, using the same method as in the practice sessions but starting from other locations. |
| (Hørlyck et al. 2021) | Computer in a virtual, non-immersive office environment. | The JEF test consists of a number of tasks that are designed to place demands on 8 subdomains of executive function as planning, prioritization, selection, creativity, adaptiveness, action-based prospective memory, time-based prospective memory and event-based prospective memory. Participants complete these tasks in a virtual, non-immersive office environment on a computer. Neuropsychological test battery: RAVLT, Trail Making A and B, verbal fluency tests, WAIS-III, letter-number sequencing and the RBANS digit span and coding tests. Functional measures: FAST and UPSA-B. Reading Test. |
| (Ilioudi et al. 2023) | Wireless, 3 degrees-of-freedom VR head-mounted display (an Oculus Go) as the VR equipment for the VR calm room intervention. The VR app used in the study was developed in the game engine "Unity3D" and was also released on the Oculus store | A variety of soothing nature environments could be tailored through (e.g. day and night scenes and dynamic weather). Breathing exercises, mindfulness programs, and relaxing music. |
| (Kim 2008) | HMD. | Individuals adjust their distance to avatars for comfortable conversation, then they measure the angle of gaze deviation from avatar's forehead, and subjects rate emotional feelings and excitement toward avatars using the Self-Assessment Manikin scale. |
| (KimJung et al. 2009) | HMD. | Individuals were presented with emotional avatars and asked to infer the reason for the avatar's emotion based on verbal and nonverbal cues. The task consisted of six experimental blocks and six control blocks, and each block had three phases: observation, inference, and response. The participants were asked to judge a short sentence that suggested a potential reason for the avatar's emotion by clicking a mouse button. |
| (KimKu et al. 2009) | HMD. | Individuals were asked to approach an avatar and stop at a distance where they felt most comfortable to have a conversation. From that distance, subjects were to say "hello" first and wait for the response of the avatar. The avatar would then talk about topics compatible with its facial emotion and request that the subjects introduce themselves. Then, subjects were to introduce themselves and say "over" to announce that the self-introduction was finished. Individuals rate emotional feelings and excitement toward avatars using the Self-Assessment Manikin scale. |
| (Kjærstad et al. 2022) | Fully immersive 360-degree equipment. | Individuals completed two different paradigms to assess their emotional regulation abilities: a traditional behavioural paradigm and a novel VR-based social scenarios test called the VERA test. In the VERA test, participants were presented with social scenarios in a fully immersive 360-degree VR environment and were instructed to either simply view the scenario or try to dampen their emotional response. The scenarios were presented in a fixed order (neutral, negative, positive), and each condition concluded with a self-rated emotion rating. The traditional behavioural paradigm involved participants down-regulating emotions to descriptions of social scenarios on a computer screen. |
| (Miskowiak et al. 2021) | CAVIR is administered on a standalone head-mounted Oculus Go 32 GB portable headset running on a 5.5 inch LCD display with a resolution 1280x1440 pixels per eye at 72hz refresh rate. The headset uses a Qualcomm Snapdragon 821 chipset as the main CPU and GPU processing unit. | CAVIR: The tasks include memorizing a list of ingredients and taking them out of the fridge, planning and selecting the order of sub-tasks involved in cooking a meal, placing correct ingredients in a pot based on a key of symbols, observing and memorizing the location of cutlery and flatware in the kitchen cupboards and drawers, and repeatedly checking the lasagne in the oven in response to a specific combination of visual and auditive cues while ignoring irrelevant stimuli. Standard neuropsychological measures: The One Touch Stocking of Cambridge, SWM test and RVP from CANTAB, 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 DART. Functional assessments: FAST (observer-based scale) UPSA-B, performance-based measure of functioning, assessing Financial and Communication skills. |
| (Mohammadi et al. 2018) | VRNT with two environments - a virtual neighbourhood and a virtual maze - to assess participants' spatial navigation skills through recalling routes. The VRNT featured 3D first-person and 2D overhead views, and participants viewed the overhead environment for 60 seconds, followed by the 3D view, locating goals marked in the overhead view (yellow circle), and recording responses and reaction times. The high-quality 3D images were produced using AutoCAD, Adobe Photoshop, and Unity 3D software. Movements controlled bia joystick. | The VRNT featured 3D first-person and 2D overhead views, and individuals viewed the overhead environment (two environments - a virtual neighbourhood and a virtual maze) for 60 seconds, followed by the 3D view, locating goals marked in the overhead view (yellow circle), recording responses and reaction times. The movements in the environment were controlled using a joystick. Neuropsychological assessment: Verbal memory: AVLT, Non-verbal memory: ROCFT-R Visuospatial memory: ROCFTD and MMSE |
| (Perra et al. 2023) | The fully immersive VR intervention was delivered using the "Oculus Go" VR view: a CE-marked device developed by Facebook Technologies in partnership with Qualcomm and Xiaomi. The Oculus Go is an all-in-one headset that contains all the components needed to provide VR experiences and does not need to be tethered to an external device to use. | The CEREBRUM app offered 52 exercises of varying difficulty, divided into three modules: Memory and Learning, Cognitive Estimates, and Attention and Working Memory. During the VR exposure, participants answered the health worker's questions while exploring the 360º scenario. The intervention consisted of 24 sessions of 45 minutes each, divided into two sessions per week over three months. Each session included a structured sequence of activities, including reception, psychoeducation, exercise psychoeducation, execution of the exercise in VR with positive and corrective feedback, post-exercise comments, and homework. |
| (Shah et al. 2015) | A lightweight, head-mounted display (model ITG-PCX3) was used to guide the relaxation practice. | VR DE-STRESS Program comprised two major components: psychoeducation and relaxation practice guided by VR-based relaxation videos. The program involved viewing relaxation videos showcasing three relaxation techniques: abdominal breathing, muscle relaxation, and guided imagery. The relaxation videos contained visual presentations, soothing auditory instructions (with female voice in English), and relaxing music. The participants were asked to practice the relaxation techniques while viewing the videos. |

AVLT: Auditory-Verbal Learning Test; CANTAB: Cambridge Neuropsychological Test Automated Battery; CAVE: Cave Automatic Virtual Environment; CAVIR: Cognition Assessment in Virtual Reality; HMD: Head-mounted displays; JEF: Jansari assessment of Executive Functions. FAST: Functioning Assessment Short Test; MMSE: Mini-Mental State Examination; RAVLT: Rey Auditory Verbal Learning Test; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status Effort Scale; ROCFT-R: Rey-Osterrieth Complex Figure Test (ROCFT-R); ROCFTD: Rey-Osterrieth Complex Figure Test; RVP: Rapid Visual Information Processing; SWM: Spatial Working Memory; UPSA-B: Performance-based Skills Assessment; VRNT: Virtual Reality Navigation Test; WAIS-III: Wechsler Adult Intelligence Scale.

**eTable 4:** Quality assessment on the included studies

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **first author, year** | **Selection bias** | **Study design** | **Confounders** | **Blinding** | **Data collection method** | **Withdrawals and dropouts** | **Global** |
| (Gould et al. 2007)\* | Moderate | Moderate | Strong | Moderate | Strong | Weak | **Moderate** |
| (Hørlyck et al. 2021)\* | Moderate | Moderate | Moderate | Moderate | Strong | Weak | **Moderate** |
| (Ilioudi et al. 2023)\* | Strong | Moderate | Moderate | Moderate | Strong | Weak | **Strong** |
| (Kim 2008)\* | Moderate | Moderate | Strong | Moderate | Strong | Weak | **Moderate** |
| (KimJung et al. 2009)\* | Moderate | Moderate | Moderate | Moderate | Strong | Weak | **Moderate** |
| (KimKu et al. 2009) | Moderate | Moderate | Moderate | Moderate | Strong | Weak | **Moderate** |
| (Kjærstad et al. 2022)\* | Moderate | Moderate | Strong | Moderate | Strong | Weak | **Moderate** |
| (Miskowiak et al. 2021)\* | Moderate | Moderate | Moderate | Moderate | Strong | Weak | **Moderate** |
| (Mohammadi et al. 2018)\* | Moderate | Moderate | Moderate | Moderate | Strong | Weak | **Moderate** |
| (Perra et al. 2023)\* | Moderate | Strong | Strong | Moderate | Strong | Moderate | **Moderate** |
| (Shah et al. 2015)\* | Strong | Moderate | Strong | Moderate | Strong | Strong | **Strong** |

**REFERENCES**

**Gould NF, Holmes MK, Fantie BD, Luckenbaugh DA, Pine DS, Gould TD, Burgess N, Manji HK and Zarate CA** (2007) Performance on a virtual reality spatial memory navigation task in depressed patients. *Am J Psychiatry* **164**(3)**,** 516-519. <https://doi.org/10.1176/ajp.2007.164.3.516>.

**Hørlyck LD, Obenhausen K, Jansari A, Ullum H and Miskowiak KW** (2021) Virtual reality assessment of daily life executive functions in mood disorders: associations with neuropsychological and functional measures. *J Affect Disord* **280**(Pt A)**,** 478-487. <https://doi.org/10.1016/j.jad.2020.11.084>.

**Ilioudi M, Lindner P, Ali L, Wallström S, Thunström AO, Ioannou M, Anving N, Johansson V, Hamilton W, Falk Ö and Steingrimsson S** (2023) Physical Versus Virtual Reality-Based Calm Rooms for Psychiatric Inpatients: Quasi-Randomized Trial. *J Med Internet Res* **25,** e42365. <https://doi.org/10.2196/42365>.

**Kim E** (2008) Characteristics of Social Behavior of Patients with Bipolar Disorder Measured by Proximal Spacing in Immersive Virtual Reality Environment Yonsei University.

**Kim E, Jung YC, Ku J, Kim JJ, Lee H, Kim SY, Kim SI and Cho HS** (2009) Reduced activation in the mirror neuron system during a virtual social cognition task in euthymic bipolar disorder. *Prog Neuropsychopharmacol Biol Psychiatry* **33**(8)**,** 1409-1416. <https://doi.org/10.1016/j.pnpbp.2009.07.019>.

**Kim E, Ku J, Kim JJ, Lee H, Han K, Kim SI and Cho HS** (2009) Nonverbal social behaviors of patients with bipolar mania during interactions with virtual humans. *J Nerv Ment Dis* **197**(6)**,** 412-418. <https://doi.org/10.1097/NMD.0b013e3181a61c3d>.

**Kjærstad HL, Hellum KS, Haslum NH, Lopes MN, Noer TS, Kessing LV and Miskowiak KW** (2022) Assessment of the validity and feasibility of a novel virtual reality test of emotion regulation in patients with bipolar disorder and their unaffected relatives. *J Affect Disord* **318,** 217-223. <https://doi.org/10.1016/j.jad.2022.09.004>.

**Miskowiak KW, Jespersen AE, Kessing LV, Aggestrup AS, Glenthøj LB, Nordentoft M, Ott CV and Lumbye A** (2021) 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. *J Psychiatr Res* **145,** 182-189. <https://doi.org/10.1016/j.jpsychires.2021.12.002>.

**Mohammadi A, Hesami E, Kargar M and Shams J** (2018) Detecting allocentric and egocentric navigation deficits in patients with schizophrenia and bipolar disorder using virtual reality. *Neuropsychol Rehabil* **28**(3)**,** 398-415. <https://doi.org/10.1080/09602011.2017.1369888>.

**Perra A, Galetti A, Zaccheddu R, Locci A, Piludu F, Preti A, Primavera D, Di Natale L, Nardi AE, Kurotshka PK, Cossu G, Sancassiani F, Stella G, De Lorenzo V, Zreik T and Carta MG** (2023) A Recovery-Oriented Program for People with Bipolar Disorder through Virtual Reality-Based Cognitive Remediation: Results of a Feasibility Randomized Clinical Trial. *J Clin Med* **12**(6). <https://doi.org/10.3390/jcm12062142>.

**Shah LB, Torres S, Kannusamy P, Chng CM, He HG and Klainin-Yobas P** (2015) Efficacy of the virtual reality-based stress management program on stress-related variables in people with mood disorders: the feasibility study. *Arch Psychiatr Nurs* **29**(1)**,** 6-13. <https://doi.org/10.1016/j.apnu.2014.09.003>.