Virtual Reality in mental health assessment: acceptability, usability and insights into cybersickness levels of a novel VR environment for the evaluation of depressive symptoms

Sara Sutori, Emma Eliasson, Francesca Mura, Victor Ortiz, Vincenzo Catrambone, Gergö Hadlaczky, Ivo Todorov, Antonio Luca Alfeo, Valentina Cardi, Mario G.C.A. Cimino, Giovanna Mioni, Mariano Alcañiz Raya, Gaetano Valenza, Vladimir Carli, Claudio Gentili

Submitted to: Journal of Medical Internet Research
on: November 08, 2024

Table of Contents

Virtual Reality in mental health assessment: acceptability, usability and insights into cybersickness levels of a novel VR environment for the evaluation of depressive symptoms

Sara Sutori[1*]; Emma Eliasson[1*]; Francesca Mura[2]; Victor Ortiz[3]; Vincenzo Catrambone[4]; Gergö Hadlaczky[1]; Ivo Todorov[1]; Antonio Luca Alfeo[4]; Valentina Cardi[5]; Mario G.C.A. Cimino[4]; Giovanna Mioni[5]; Mariano Alcañiz Raya[3]; Gaetano Valenza[4]; Vladimir Carli[1*]; Claudio Gentili[5*]

[1]National Centre for Suicide Research and Prevention (NASP) Department of Learning, Informatics, Management and Ethics (LIME) Karolinska Institutet Stockholm SE
[2]Padova Neuroscience Center (PNC) University of Padua Padua IT
[3]Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano Universitat Politècnica de València Ciudad Politécnica de la Innovación València ES
[4]Research Center E.Piaggio Department of Information Engineering University of Pisa Pisa IT
[5]Department of General Psychology University of Padua Padua IT
[*]these authors contributed equally

Corresponding Author:
Emma Eliasson
National Centre for Suicide Research and Prevention (NASP)
Department of Learning, Informatics, Management and Ethics (LIME)
Karolinska Institutet
Granits väg 4
Stockholm
SE

Abstract

Background: There is a clear need for enhanced mental health assessment, depressive symptom evaluation being no exception. A promising approach to this aim is utilising Virtual Reality (VR), that entails the potential of adding a wider set of assessment domains with enhanced ecological validity. However, whilst several studies have utilised VR for both diagnostic and treatment purposes, its acceptance, in particular how exposure to virtual environments affect populations with psychiatric conditions remains unknown.

Objective: The current study reports on the acceptability, usability and cybersickness levels of a pilot VR-environment designed for the purpose of differentiating between individuals with depressive symptoms.

Methods: The study, conducted in Italy, included 50 healthy controls and 50 individuals with mild to moderate depressive symptoms and employed an observational design with circa 30-minute VR exposure followed by a self-report questionnaire battery.

Results: Results indicate that the majority found VR acceptable for the purposes of mental health screening and treatment. However, for diagnostics, there was a clear preference for VR to be used by mental health professionals as a supplementary tool, as opposed to a standalone solution. In practice, following exposure to the pilot VR-environment, generally good levels of acceptability and usability were reported, but areas in need of improvement were identified (such as self-efficacy). Self-reported cybersickness levels were considerably higher among those with depressive symptoms. This finding raises questions about the potential interplay between underlying somatic symptoms of depression and VR-induced cybersickness and calls for more attention from the scientific community both in terms of methodology as well as clinical and theoretical implications.

Conclusions: Conclusively, user support indicates a potential for VR to aid mental health assessment, but further research is needed to understand how exposure to virtual environments might affect populations with psychiatric symptoms.

(JMIR Preprints 08/11/2024:68132)
DOI: https://doi.org/10.2196/preprints.68132

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Cover page

Paper type

Original paper

Title

Virtual Reality in mental health assessment: acceptability, usability and insights into cybersickness levels of a novel VR environment for the evaluation of depressive symptoms

Authors

Sara Sutori[1]*, Emma Eliasson[1]*†, Francesca Mura[2], Victor Ortiz[3] Vincenzo Catrambone[4], Gergö Hadlaczky[1], Ivo Todorov[1], Antonio Luca Alfeo[4], Valentina Cardi[5], Mario G.C.A. Cimino[4], Giovanna Mioni[5], Mariano Alcañiz Raya[3], Gaetano Valenza[4], Vladimir Carli[1]*, Claudio Gentili[5]*

1 National Centre for Suicide Research and Prevention (NASP), Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden
2 Padova Neuroscience Center (PNC), University of Padua, Padua, Italy
3 Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, Universitat Politècnica de València. Ciudad Politécnica de la Innovación, València, Spain
4 Research Center E.Piaggio, Department of Information Engineering, University of Pisa, Pisa, Italy
5 Department of General Psychology, University of Padua, Padua, Italy

* These authors contributed equally to this work.
† Corresponding author: Emma Eliasson
Address: Granits väg 4, 171 65 Solna, Sweden
Phone: +46 70-140 72 95
Fax: 08-31 11 01
Email: emma.eliasson@ki.se

Author contributions

Conceptualization: GV, VlC, CG, MAR; Methodology: EE, SS, FM, VO, GH, IT, VaC, VlC, CG; Software: VO; Validation: VlC, CG; Formal analysis: SS, EE, FM; Investigation: FM; Resources: GV, VlC, CG, MAR; Data Curation: EE, SS, FM, VO; Writing - Original Draft: EE, SS, FM, VO; Writing - Review & Editing: GH, IT, ViC, MGCAC, GM, VaC, MAR, GV, VlC, CG; Visualization: VO, SS; Supervision: GV, VlC, CG, MAR; Project administration: VlC, CG, MAR, GV; Funding acquisition: GV, VlC, CG, MAR.

Abstract

There is a clear need for enhanced mental health assessment, depressive symptom evaluation being no exception. A promising approach to this aim is utilising Virtual Reality (VR), that entails the potential of adding a wider set of assessment domains with enhanced ecological validity. However, whilst several studies have utilised VR for both diagnostic and treatment purposes, its acceptance, in particular how exposure to virtual environments affect populations with psychiatric conditions remains unknown. The current study reports on the acceptability, usability and cybersickness levels of a pilot VR-environment designed for the purpose of differentiating between individuals with depressive symptoms. The study, conducted in Italy, included 50 healthy controls and 50 individuals with mild to moderate depressive symptoms and employed an observational design with circa 30-minute VR exposure followed by a self-report questionnaire battery. Results indicate that the majority found VR acceptable for the purposes of mental health screening and treatment. However, for diagnostics, there was a clear preference for VR to be used by mental health professionals as a supplementary tool, as opposed to a standalone solution. In practice, following exposure to the pilot VR-environment, generally good levels of acceptability and usability were reported, but areas in need of improvement were identified (such as self-efficacy). Self-reported cybersickness levels were considerably higher among those with depressive symptoms. This finding raises questions about the potential interplay between underlying somatic symptoms of depression and VR-induced cybersickness and calls for more attention from the scientific community both in terms of methodology as well as clinical and theoretical implications. Conclusively, user support indicates a potential for VR to aid mental health assessment, but further research is needed to understand how exposure to virtual environments might affect populations with psychiatric symptoms.

Keywords

Depression, virtual reality, assessment, acceptability, usability, cybersickness

Abbreviations

AOI	Area of Interest
CEI-II	Curiosity and Explorations Inventory
DASS	Depression, Anxiety, Stress Scales
DSM	Diagnostic and Statistical Manual of Mental Disorders
PANAS-SF	Positive and Negative Affective Schedule – Short Form
PHQ-9	Patient Health Questionnaire
POM	Proportional Odds Model
RVIP	Rapid Visual Information Processing
SSQ	Simulator Sickness Questionnaire
SUS	System Usability Scale
TFA	Theoretical Framework of Acceptability
TMT	Trail Making Test
VE	virtual environment
VR	virtual reality
WCST	Wisconsin Card Sorting Test

Introduction

In 2019, an estimated 970 million people worldwide, or 12.6% of the global population, were living with a mental disorder, with anxiety (301 million) and depression (280 million) being the most prevalent [1]. These conditions contribute substantially to the global burden of disease, with depression being one of the leading causes of disability worldwide [2]. Despite considerable efforts to address this issue, there has been no substantial reduction in the burden of mental disorders since 1990 [1].

Depressive disorders (hereafter referred to as depression) present significant challenges in research, amongst other reasons due to the diagnostic uncertainty [3], characterized by both high heterogeneity and low reliability [4]. The persistence of these challenges over different iterations of the Diagnostic and Statistical Manual of Mental Disorders [DSM] [4] underscores the need to advance clinical practice, which currently predominantly relies on self-report questionnaires and clinical interviews [5].

In light of the abovementioned challenges, the idea has been brought forward to include a wider battery of measures that covers a broader spectrum of assessment domains and is not reliant solely on verbal self-report [6,7]. Such measures include speech, text, and facial expression analysis [8]; genomics, transcriptomics, proteomics, metabolomics and imaging [9]; behavioral and physiological variables collected via wearable sensors [10] or Virtual Reality systems with the potential to combine multiple domains of information [7]. However, while efforts are being dedicated to the testing of various measures to enhance assessment, less attention has been paid to the evaluation of such novel approaches from a user perspective.

Beyond diagnostic accuracy, evaluating acceptability and usability of diagnostic tools is crucial, particularly when involving individuals with mental health conditions, like depression. In such cases, where patients might face additional treatment barriers, including motivational difficulties, stigma, delayed help-seeking, [11] or poor treatment adherence [12], these factors play a crucial role in the successful implementation of new diagnostic and treatment technologies.

Sekhon and colleagues [13] define *acceptability* as a “multi-faceted construct that reflects the extent to which people delivering or receiving a healthcare intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention”. On the other hand, *usability* has various definitions and approaches (for an overview see [14]), but here we will refer to the extent to which the Virtual Reality (VR) system can be used with ease for the purpose of assessing depression symptom severity.

This study introduces a novel approach aimed at enhancing the quality - including the breadth and ecological validity - of mental health evaluations. A VR system and environment was developed to capture physiological, behavioral, and cognitive data associated with depression as well as symptom severity. Given the innovative nature of this approach, it is fundamental to consider the patient perspective as well. Therefore, the objective of this study was to evaluate the acceptability, usability and cybersickness levels of the pilot virtual environment.

Aim

The aim of the study was to assess the acceptability, usability and cybersickness levels of a pilot VR-environment designed to aid and enrich the assessment of depressive symptoms.

Methods

Sample

A total of 266 individuals were screened at the University of Padua, Italy, with continuous enrollment between November 2022 and November 2023. Recruitment was completed when the predefined target sample size of 100 was reached, including 50 participants with depressive symptoms and 50 participants classed as healthy controls. Participants were aged 18-35 years, fulfilled all inclusion criteria, and provided written informed consent. Compensation included 25 EUR. Further demographic and clinical characteristics of the sample are outlined in Table 1.

Table 1. Demographic and clinical characteristics of study participants by study groups.

	Depressed†	**Controls**	**Between group difference (*r*)††**
n	50	50	
Sex [males/females]	10/40	10/40	
Age [years]	23.0 (1.86)	23.3 (1.52)	0.13
Education [years]	16.1 (1.40)	16.6 (0.78)	0.14
Hand dominance [left/right]	0/50	3/47	
Sight correction [none/lenses/glasses]	21/10/19	26/10/14	
Height [m]	1.67 (0.07)	1.66 (0.07)	0.07
Weight [kg]	62.7 (10.60)	59.6 (7.85)	0.12
Body Mass Index	22.4 (3.00)	32.5 (2.06)	0.13
Habitual sleep [hours / night]	7.0 (0.98)	7.3 (0.88)	0.20
Habitual smoker [yes/no]	19/31	20/30	
Habitual consumption of alcohol [yes/no]	11/39	16/34	
Habitual user of psychoactive drugs [yes/no]	11/39	4/46	
PHQ-9	11.3 (1.92)	3.2 (1.48)	0.87***
DASS-D	15.2 (7.40)	3.5 (3.44)	0.77***
DASS-A	9.6 (6.62)	2.4 (2.89)	0.61***
DASS-S	17.8 (6.73)	7.7 (5.08)	0.66***
PANAS-Pos *before VR*	10.4 (3.53)	12.2 (2.76)	0.27**
PANAS-Pos *after VR*	8.9 (4.36)	11.8 (3.61)	0.36***
PANAS-Neg *before VR*	3.2 (2.98)	1.9 (2.22)	0.24*
PANAS-Neg *after VR*	3.2 (3.25)	1.7 (2.24)	0.24*

Notes: Sex, hand dominance, sight corrections, being habitual smoker, alcohol or psychoactive drug consumer are all given as frequency counts. Other variables are given as mean and (standard deviation).

† The ‘Depressed’ categorization is based on current depressive symptom severity assessment via the PHQ-9 questionnaire (score equal to 9 or above) and not the presence of a clinical diagnosis.

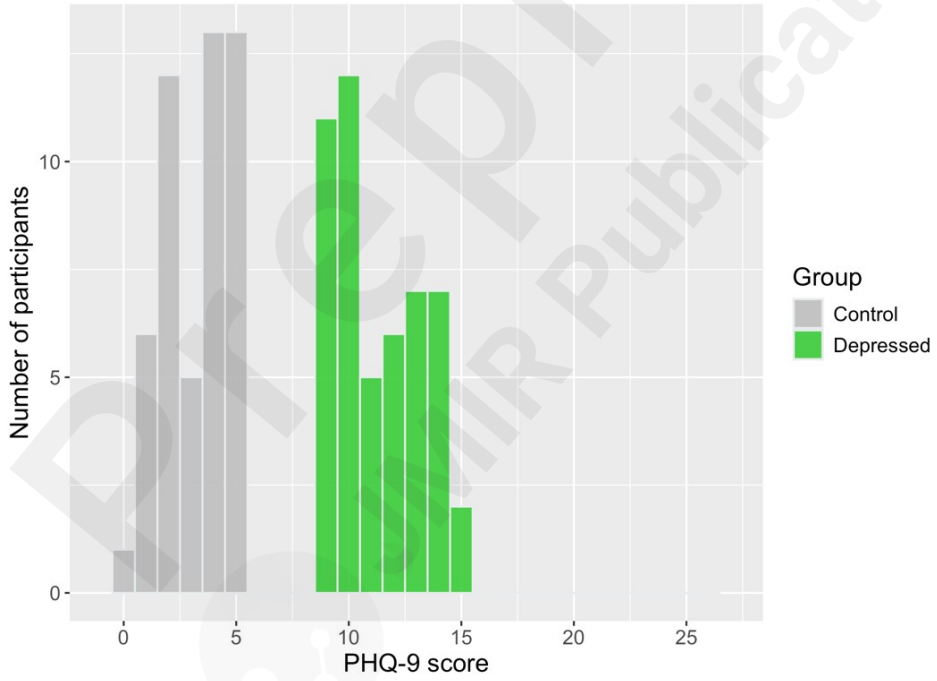
†† Group differences are calculated via two-tailed Mann Whitney U tests. Significance: * $P<.05$; ** $P<.01$; *** $P<.001$

Inclusion and exclusion criteria

The screening for inclusion involved the assessment of current depressive symptom severity using the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001). Participants scoring 9 or above - constituting the upper limit for mild depressive symptoms [15] - were assigned to the case group (hereafter denoted as ‘depressed’), while those scoring 5 or below were registered for the control group (see Fig 1 for score distributions).

For inclusion in either group, participants had to be free from any condition that would impair their ability to interact with the VR environment (e.g., visual impairment without the possibility of correction via lenses) or compromised with their ability to provide written informed consent (e.g. intellectual disability). Additional inclusion criteria for the ‘depressed’ group included the lack of diagnosis of any other psychiatric disorders than depressive or anxiety disorders, and the stability of treatment over the last 4 weeks if they received psychological or pharmacological treatment at the time of assessment. Participants in the control group had to be free from any previous or current psychiatric disorders.

Figure 1. Distribution of participants’ scores on the Patient Health Questionnaire (PHQ-9) assessing current depressive symptom severity.



Notes. The PHQ-9 scores range from 0 to 27; where 0–4 indicates no depressive symptoms, 5–9 means mild depressive symptoms, 10–14 corresponds to moderate depressive symptoms, 15–19 shows moderately-severe depressive symptoms, and 20–27 signals severe depressive symptoms [15].
‘Depressed’ categorization is solely dependent on the PHQ-9 score and does not indicate the presence of a clinical diagnosis.

Study design

The study followed an observational design, where all participants completed the same questionnaire batteries pre- and post-VR (detailed below) and followed the same VR-exposure protocol.

Study procedure

After providing informed consent and completing the pre-VR battery questionnaires (detailed below), all participants were seated in a room equipped with the VR system. Following a headset and eye-tracking calibration, a tutorial was completed (aimed to help the understanding of how to move around and interact with the environment). Next, participants were free to explore the virtual reality environment on their own terms but in the presence of a study administrator. The study session was concluded once the post-VR battery was completed.

Technological specifications

The virtual reality system was developed using the HP Reverb G2 Omnicept Edition headset, which includes a variety of biometric sensors such as eye tracking, a facial camera (lower), a heart rate sensor, and sensors for brain activity, while maintaining the core features of the standard version. These features include a resolution of 2160 x 2160 px per eye, a 114° field of view, a 90 Hz refresh rate, and integrated audio.

Environment development

The virtual environment was developed entirely in Unity 2020.3.39 LTS, utilizing the HP Omnicept SDK and OpenXR to maximize compatibility and platform versatility. Several development processes were employed to ensure a smooth and efficient environment, including occlusion culling, foveated rendering, baked lighting, and asset optimization.

The scenario consists of four rooms (detailed later), each with distinct lighting parameters, as well as an additional room that can be explored through a keyhole in a door. Volume post-processing was used to modify the lighting without affecting performance, and the keyhole effect was simulated using camera layers.

Within each room, various stimuli and interactive elements are present, such as sound boxes, a chalkboard, interactable objects, musical buttons, and photos that can be picked up for closer inspection. Every interaction between the subject and the environment is recorded, including eye-tracking data, particularly in relation to elements designated as Areas of Interest (AOIs) or with which the subject can interact (e.g., ENTERING_ROOM, LEAVING_ROOM, TAKE, DROP, TRIGGER, INTERACT, etc.).

Administrator control

To manage the experiment, the study administrator is provided with an interface that allows them to observe the subject's actions in real-time. The interface also offers various options to ensure that the experimental session proceeds smoothly. This control system is facilitated by an additional camera mounted on the subject, enabling the administrator to monitor the experiment's progression effectively.

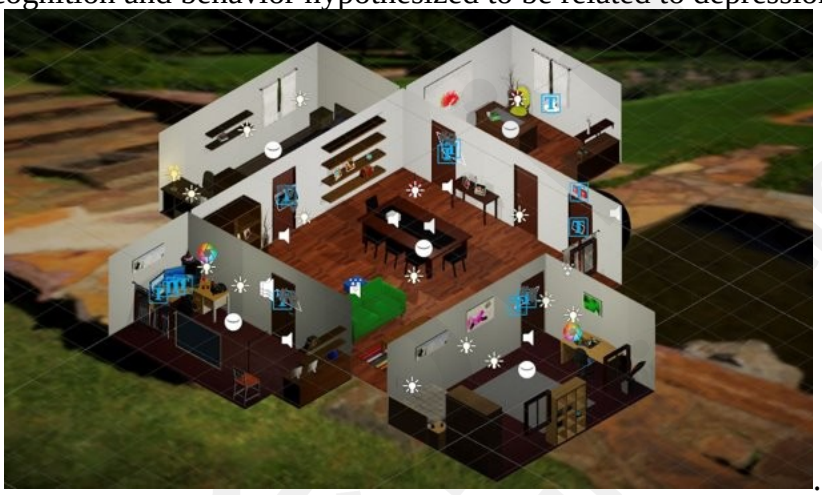
Virtual environment

The goal of the pilot VR environment is to differentiate between healthy controls and those with depressive symptoms based on cognitive, behavioral and physiological data. To this aim, the environment was designed to elicit certain behaviors and record measures, some of which were previously shown, others are hypothesized to be connected to depression. The domains considered include cognition (attention, working memory, processing speed, executive functioning, cognitive flexibility), metacognition, persistence/grit, curiosity, as well as behavioral and attentional biases

towards negative stimuli assessed via interactions in the environment, eye-tracking, speed and movement measurements. Physiological data included skin conductance and heart-rate variability.

The environment resembles a multi-room family home (see Fig 2). The exposure started with a tutorial introducing participants to controllers and how to move around in the environment and interact with objects. Following the completion of the tutorial, participants were free to explore the environment as they wish, but do so, to move between rooms of the environment, they must solve cognitive tasks to open doors (four in total: Rapid Visual Information Processing [RVIP]; N-back [2-back]; Trail Making Test - Parts A & B [TMT]; Wisconsin Card Sorting Test [WCST]). The exposure ends on the participant's own terms, but the exit door only opens once all cognitive tasks have been completed.

Figure 2. The visual illustration of the virtual reality environment designed for the assessment of cognition and behavior hypothesized to be related to depression



Tools and questionnaires

Pre-VR questionnaires
The pre-VR battery included a background questionnaire (demographic data, diagnosis, medication); the Patient Health Questionnaire (PHQ-9; [15]) to assess current depressive symptom severity; the Depression, Anxiety, Stress Scale (DASS-21; [16]) to assess current depressive, anxiety and stress symptom severity; the Positive and Negative Affect Schedule (PANAS-SF; [17]) to evaluate emotional state; as well as the Curiosity and Explorations Inventory-II (CEI-II; [18]).

Post-VR questionnaires
The post-VR battery repeated the PANAS-SF questionnaires, as well as questions regarding, acceptability, usability and cybersickness, described in more detail below.

Acceptability. The construct of acceptability was assessed based on the Theoretical Framework of Acceptability (TFA; [13,19]). The acceptability of the concept and the acceptability of the pilot system were evaluated separately. The acceptability of the *concept*, i.e. the use of VR for the purpose of mental health screening, diagnosis and treatment (definitions were provided) was assessed via self-report Likert items. Scores ranged from 1 to 5, indicating answers from "Strongly disagree" to "Strongly agree" on statements about VR being acceptable (see Supplementary Material 1 for details). As the primary purpose of the pilot system tested here was the assessment of depressive symptoms, the diagnostic use case was investigated via two additional questions: (1) willingness to engage with the system to inform diagnosis made by a mental health professional, and (2)

willingness to engage with the system to receive a diagnosis without extra input from a mental health professional.

The dimensions of acceptability concerning the pilot system were assessed within the TFA and are focused on General acceptability, Affective attitude, Perceived effectiveness, Intervention coherence, Self-efficacy, Burden, Ethicality and Opportunity cost. The construct of Affective attitude is further broken down to two questions: (1) to what extent does a participant like or dislike the system; (2) how comfortable or uncomfortable they find the system to be. For the ease of interpretability, the rating scale was unionized across all constructs, where '1' represents the "undesirable" rating (e.g. low levels of Comfort; or the need for high levels of effort indicated for Burden), and '5' reflects the "desirable" rating (e.g. high levels of Comfort; or low levels of effort reported under Burden). For the specific items used to assess each of these constructs please visit Supplementary Material 1.

Usability
Usability was assessed by the System Usability Scale (SUS; [20]). The scale is composed of 10 items rated on a Likert scale from 1 to 5, from 'strongly disagree' to 'strongly agree'. The total score ranges from 0 to a 100, with higher scores indicating higher usability ratings. There are multiple frameworks to interpret the SUS score (range, adjective, grade; see [21] for example), but the primary threshold here will be the average score for the experience to be rated as 'OK/satisfactory', which is 50.9 [22].

Cybersickness
To assess levels of cybersickness following VR-exposure the Simulator Sickness Questionnaire (SSQ) was used [23]. The questionnaire contains 16 symptoms which are rated on a scale of subjective severity: 0 (none), 1 (slight), 2 (moderate) 3 (severe). Beyond the total score, one can attain separate scores for the Nausea, Oculomotor and Disorientation subscales.
However, as the SSQ has been increasingly criticized, including the original thresholds for the interpretation of the severity levels [24–27], comparison of the pilot system results was made to the literature average.

Analysis

All analyses were carried out in R supplemented by R Studio ([28]; Version: 2023.12.1.402). Multiple packages were used to supplement analysis and visualization, these include but are not limited to: dplyr [29], vioplot [30], ggplot2 [31], psych [32], rstatix [33], ggeffects [34], ordinal [35], VGAM [36]. For the script, its output and full list of packages, please see Supplementary Material 2.

Simple within-group comparisons were made via Wilcoxon signed rank tests, while between-group comparisons via Mann-Whitney U tests to account for the ordinal-type outcome measures, such as Likert scales, as well as any deviation from the parametric requirements of the dependent and independent sample t-tests. The nominal α level was .05.

Models with multiple predictors were analyzed via multivariate linear regressions when the outcome variable was linear in nature (such as the SUS or SSQ scores) and via proportional odds Models (POM) for ordinal type outcome measures (such as TFA constructs) depending. In some cases (specified under the results section), the assumption of proportional odds was not fulfilled, and thus results of these models should be interpreted with caution. All regression models reported here included age, gender and education as covariates. In cases with additional predictors (such as cybersickness severity score, or group label), these were entered into the models in addition to these

demographic variables.

Ethical approval

The study obtained ethical approval from the local ethical committees on all sites where data was collected (Italy) or analyzed (Italy and Sweden). In Italy, approval was granted by Comitato Etico della Ricerca Psicologica (AREA 17), prot. No. 4688 and in Sweden by the Etikprövningsmyndigheten (2023-00959-01).

Preregistration

The study was preregistered in the ISRCTN registry under the number ISRCTN16396369 [37].

The sole modification compared to the pre-registered protocol concerns the threshold for inclusion on current depressive symptom severity. Due to the difficulty to recruit volunteers in line with the study's timeline, the upper threshold for inclusion in the control group was increased from 4 to 5, while the lower threshold to be included in the group with depressive symptoms was decreased from 10 to 9 on the PHQ-9.

Data availability

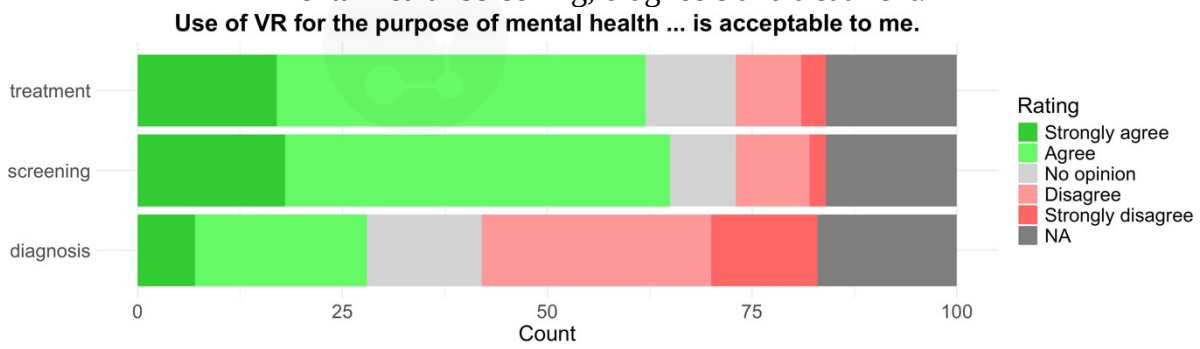
Access to the raw data will be considered on a case-by-case basis upon request. Interested parties may contact the corresponding author.

Results

Acceptability of the concept

The use of VR for mental health treatment (74%, 62/84) and screening (77%, 65/84) was endorsed by the majority of those providing an answer, but only a minority indicated support for diagnostic purposes (34%, 28/83) – see Figure 3 for a detailed breakdown. There were no significant differences between the 'depressed' and 'control' groups as investigated via Mann-Whitney U tests.

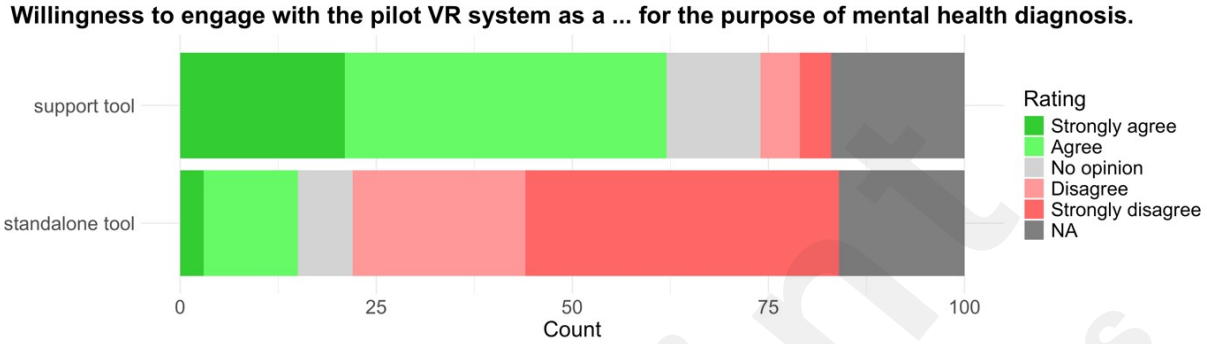
Figure 3. Distribution of scores on the acceptance of Virtual Reality technology for the purpose of mental health screening, diagnosis and treatment.



Comparing the use of VR with or without input from a mental health professional, the responses indicate higher support for the VR system to be used as support tool (75%, 62/83) compared to a standalone solution (18%, 15/84). A Wilcoxon signed rank test indicates a significant difference of

moderate to large size (V=1770, P<.001 [2-tailed], r=0.51; $mean_{support}$ 3.84, $SD_{support}$ 1.03, $mean_{standalone}$ 2.00, $SD_{standalone}$ 1.21). For the raw distribution of responses see Figure 4. Again, there was no difference between the ‘depressed’ and ‘control’ groups based on Mann-Whitney U tests.

Figure 4. Willingness to engage with the pilot VR system for the purpose to receive a diagnosis with (support tool) and without (standalone tool) input from a mental health professional.

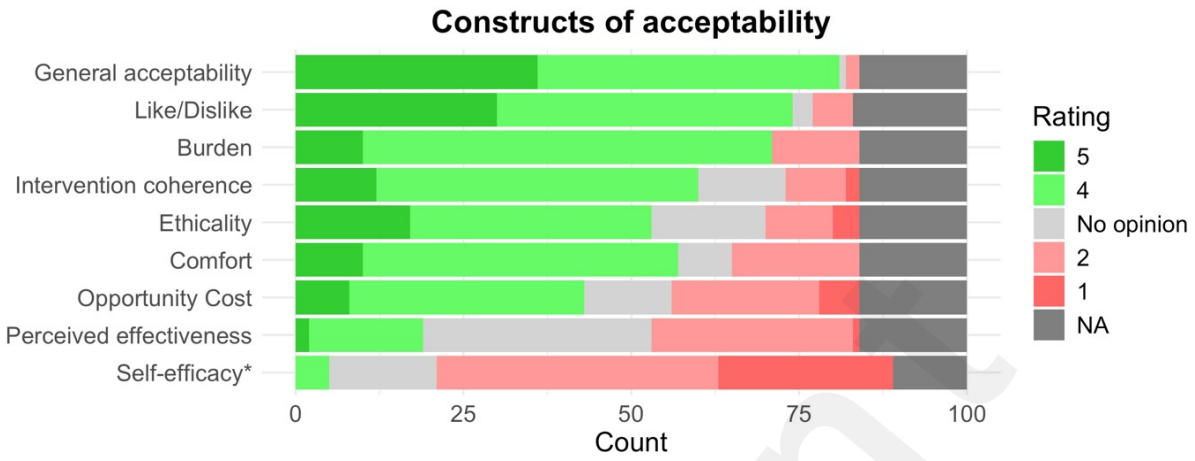


Acceptability of the pilot system

Behavioral measures of acceptability include the discontinuation of participation for any reasons, as well as the need to take a break from the study. While the possibility to discontinue participation at any point, without any consequences was stated before enrollment, all 100 volunteers completed the study. The need for a break was expressed by 1 participant due to discomfort. However, all volunteers decided to resume participation despite the possibility to opt out being re-emphasized. On average, 31 minutes were spent in the virtual environment.

Acceptability was also assessed via a self-report questionnaire specifically formulated for the project based on the Theoretical Framework of Acceptability [13,19]. The raw ratings are visualized in Figure 5 and show generally high levels of acceptability.
Majority of the participants found the system itself acceptable (96%, 81/84), liked the system (89%, 74/83), indicated that the engagement took little to no effort (85%, 71/84), found it clear how the system might help diagnose depressive disorders (71%, 60/84), indicated they find no moral or ethical consequences (63%, 53/84), found it comfortable (68%, 57/84), and expressed no concern about missing out on other alternatives when engaging with the system (51%, 43/84). The two constructs on which the undesirable ratings (1 or 2) outnumbered the desirable ones (4 or 5) were concerning the perceived effectiveness of the system in improving mental state (23%, 19/84), and self-efficacy concerning feeling confident while engaging with the system (5%, 5/99).

Figure 5. Self-reported acceptability of the pilot VR system along the constructs proposed within the Theoretical Framework of Acceptability.



Notes. Meaning of values from 1 to 5 depends on the specific construct assessed – see Supplementary Material 1. For the sake of a unified visualization, the raw values are used here, where 1 and 2 refer to undesirable ratings while 4 and 5 constitute the desirable rating on each item.
* Self-efficacy was measured with Item 9 of the System Usability Scale [20], as opposed to the other items that were specifically designed for the assessment of the pilot system based on the Theoretical Framework of Acceptability [13,19].

Multivariate ordinal regression models were utilized to explain variance in ratings of all the acceptability constructs as detailed in Table 2. Demographic variables, specifically gender was a significant predictor of general acceptability and subjective burden. Women reported considerably lower levels of acceptability (acceptability: $mean_{men}$ 4.65, SD_{men} 0.49; $mean_{women}$ 4.30, SD_{women} 0.65) and reported lower ratings on burden as well, meaning a considerable higher effort was required from their part to engage with the system ($mean_{men}$ 4.29, SD_{men} 0.50; $mean_{women}$ 3.69, SD_{women} 0.87). Education did not reach significance for any of the TFA constructs, and age was marginally significant only when used to explain whether participants find moral or ethical consequences to engaging with the system.

In the next stage, group label (‘depressed’ vs. ‘control’) was entered into the same models in addition to the demographic variables. The group label was a significant predictor of comfort levels ($mean_{control}$ 3.83, $SD_{control}$ 0.78; $mean_{depressed}$ 3.34, $SD_{depressed}$ 1.08) and marginally significant for the like/ dislike rating ($mean_{control}$ 4.38, $SD_{control}$ 0.59; $mean_{depressed}$ 4.00, $SD_{depressed}$ 0.95) - meaning that those in the ‘depressed’ group reported lower comfort levels and liked the system to a lesser extent.

Last, cybersickness severity (total SSQ score) was entered into models explaining all TFA constructs, in addition to the demographic variables. Cybersickness was a significant predictor for comfort, burden, opportunity cost and perceived effectiveness and marginally significant for self-efficacy (see Table 2). All showed an inverse relationship, whereas the higher the cybersickness symptom severity, the lower acceptability scores were reported on all TFA dimensions.

Table 2. Estimates (in log odds) of predictors from three different models analyzed via multivariate ordinal regressions to explain constructs investigated within the Theoretical Framework of Acceptability.

Constructs	Model 1 – Demographic variables			Model 2	Model 3
	Age	Education	Gender	'Depressed'§ vs. 'Control' group (+ demographic)	Cybersickness severity (+ demographic)
General acceptability	0.18	-0.53	**-1.33***	-0.26	-0.00
Affective attitude					
Like/dislike	-0.30	0.16	-0.46	-0.89	-0.01
Comfort	-0.04	0.12	-0.72	**-0.90***	**-0.03****
Burden	0.14	-0.19	**-1.87****	-0.27‡	**-0.04*****
Intervention coherence	0.09	-0.19	-0.32	-0.57	-0.00
Ethicality	0.32	-0.17	0.24	0.17	-0.01
Opportunity cost	-0.08‡	0.02‡	-0.50‡	0.34‡	**-0.02**‡
Perceived effectiveness	-0.12	-0.05	-0.76	-0.55	**-0.02***
Self-efficacy†	0.06	0.03	-0.87	0.04	-0.01

Notes: * *P*<.05; ** *P*<.01; *** *P*<.001

† Self-efficacy was measured with Item 9 of the System Usability Scale [20], as opposed to the other items that were specifically designed for the assessment of the pilot system based on the Theoretical Framework of Acceptability [13,19].

‡ The assumption of proportional odds not was not fulfilled for the specific model. Results should be interpreted with caution.

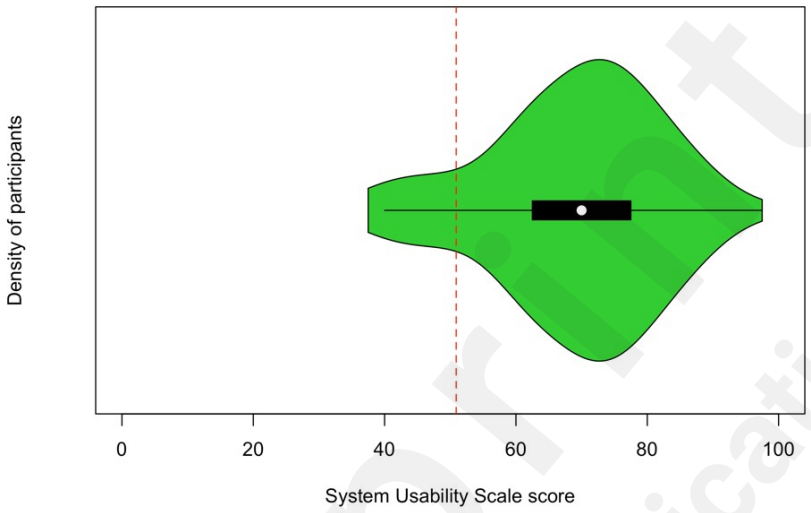
§ The 'Depressed' categorization is based on current depressive symptom severity assessment via the PHQ-9 questionnaire (score equal to 9 or above) and not the presence of a clinical diagnosis.

Usability

Descriptives

Results from 99 participants show average ratings of usability, with a mean score of 69 (SD 12.86), and a range of 37.5 – 97.5 on the System Usability Scale (SUS; [20]; see Fig 6). 88% (87/99) of participants provided a rating of 50.9 or above, which is considered to be the average of an OK/satisfactory experience [22].

Figure 6. Density plot of usability ratings from all participants.



Note. Dashed line represents the average score of when an experience is considered OK/satisfactory [22].

Analysis

A multiple linear regression analysis was conducted to predict usability ratings based on participants' age, education, gender, and group (‘depressed’ vs ’control’). The regression model was not statistically significant ($F_{4,94}$=1.23, *P*=.31), explaining only about 5% of the variance in SUS scores (R^2=0.05, adjusted R^2=0.01). The strongest predictor was gender, but this only showed a marginal effect (β=-5.79, 95% CI -12.38 to 0.80, *P*=.08) with men reporting higher usability ratings than women.

A second model was tested examining the relationship between usability scores and the predictors of age, education, gender, and total cybersickness severity. The model was found to be significant in this case ($F_{4,94}$=2.97, *P*=.02), accounting for 11.2% of the variance in SUS scores (R^2=0.11, adjusted R^2=0.07). Cybersickness severity (SSQ scores) emerged as the only significant predictor (β=-0.12, 95% CI -0.21 to -0.03, *P*=.007), indicating that participants with higher SSQ scores tended to have lower SUS scores.

Cybersickness

Descriptives

The overall mean SSQ score combining all subscales was 30.49 (SD 28.03), with individual scores ranging from 0 to 119.68. This is comparable to the literature average of 28.00 reported by Saredakis et al [38].

For the SSQ subscales, the results of the total sample reflect severity levels not (significantly) different from the literature average [38] when it comes to nausea and disorientation levels, but the

severity of oculomotor disturbances was considerably higher in this study, as shown in Table 3. However, when the comparison was split by study groups, it became apparent that it was only the 'depressed' group that deviated significantly from the literature average. The 'control' group showed near average levels on all three subscales. The distributions of raw scores along the total and all subscales are illustrated in Figure 7 and analyzed for between group differences in the section below.

Table 3. Comparison of mean levels (SD) of cybersickness severity [SSQ score] of the total sample, 'depressed', and 'control' subsamples with literature averages [38].

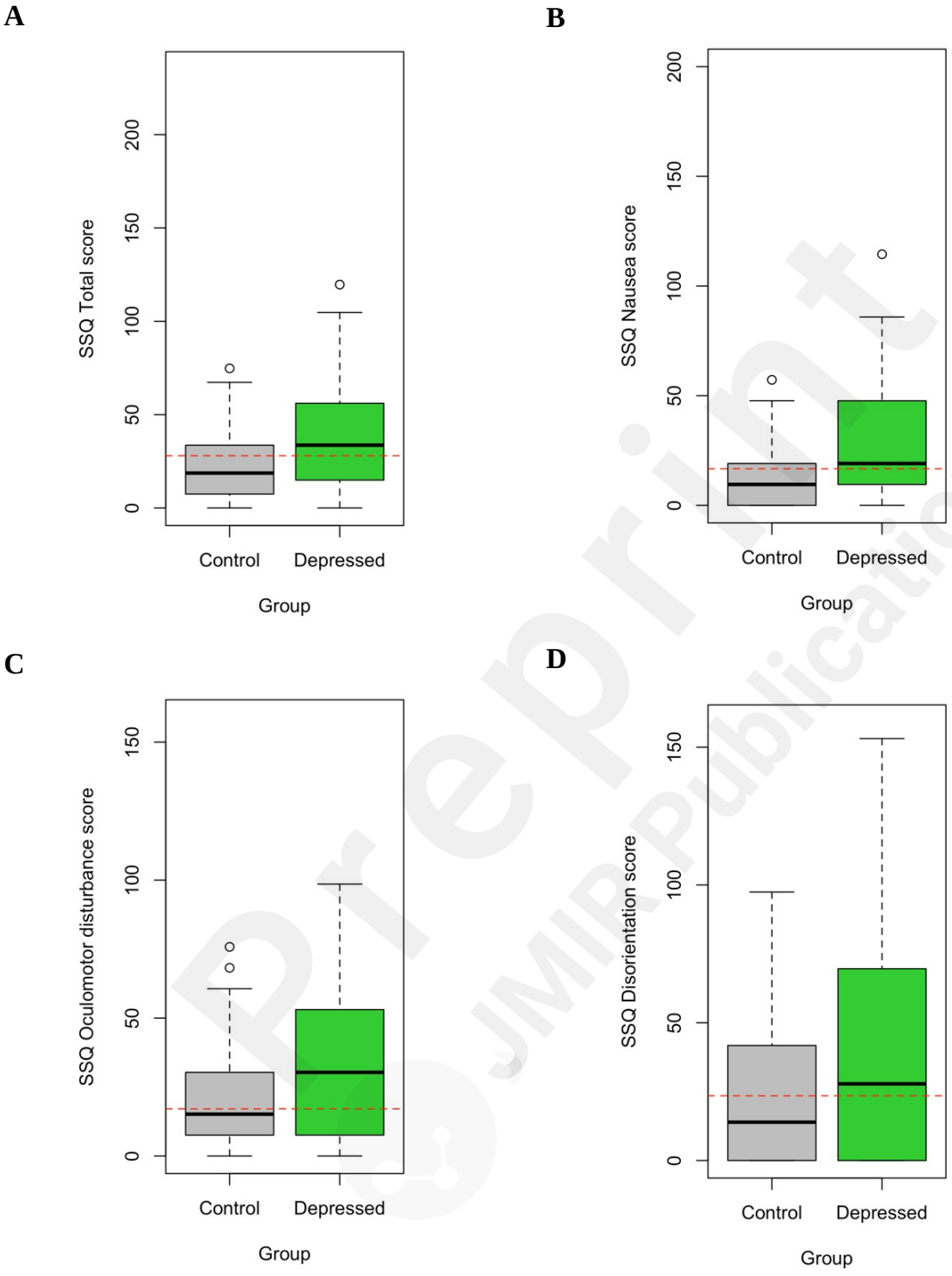
SSQ Scale	Literature average‡	Total sample	'Depressed'†	'Control'
Total	28.00	30.49 (28.03)	39.39 (32.89)	21.77 (18.84)*
Nausea	16.72	21.59 (21.96)	29.40 (24.39)**	13.93 (16.15)
Oculomotor disturbances	17.09	26.72 (24.63)**	33.57 (27.50)***	20.01 (19.48)
Disorientation	23.50	32.90 (39.84)	42.04 (48.34)	23.94 (26.83)

Notes. * *P*.05; ** *P*<.01; *** *P*<.001

† 'Depressed' categorization is solely dependent on the PHQ-9 score and does not indicate the presence of a clinical diagnosis.

‡ Literature averages are taken from Saredakis et al. [38] based on a meta-analysis of 55 articles on c. 3000 subjects.

Figure 7. Boxplot of self-reported cybersickness severity scores of participants per group on the Simulator Sickness Questionnaire (A), and its three subscales of Nausea (B), Oculomotor disturbances (C) and Disorientation (D).



Notes. Red lines represent the literature average of head mounted devices for each scale and subscale specifically, based on Saredakis et al. [38].

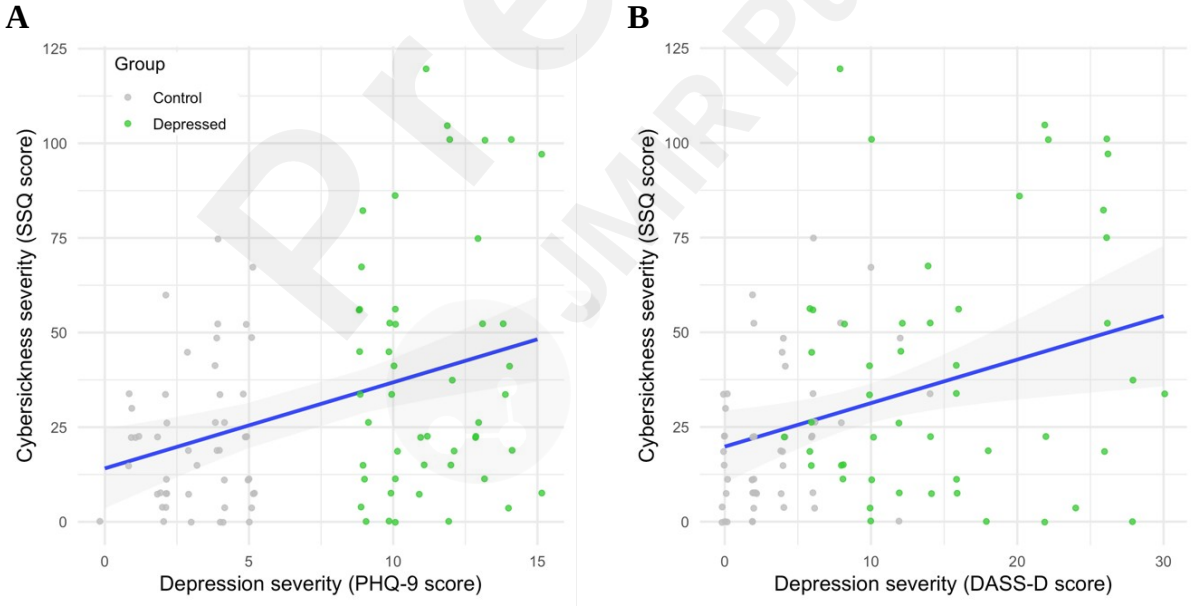
Analysis

A multiple linear regression was conducted to predict SSQ scores from age, education, gender, and group ('depressed' vs. 'control'). The model was statistically significant, $F_{4,94}$=3.20, P=.02, accounting for approximately 12% of the variance in SSQ scores (R^2=0.12, adjusted R^2=0.08). The only predictor reaching significance was the group label, (β=17.58; 95% CI 6.57 to 28.60; P=.002), indicating that participants in the 'depressed' group had considerably higher SSQ scores (mean 39.38, SD 32.89) than those in the 'control' group (mean 21.77, SD 18.84). These between-group difference remained significant for all subscales analysed separately.

When depressive symptom severity was entered as a continuous variable instead of a group label, a further 2% in variance in cybersickness severity was explained (R^2=0.14, adjusted R^2=0.11) and a linear prediction of SSQ scores along the PHQ-9 scores was attainable as visualized in Figure 8 (A).

However, due to the baseline group differences in anxiety and stress levels (Table 1) this prediction might be biased. While DASS depression was not of primary interest, a multiple linear regression model with DASS depression as an outcome was utilized, to be able to adjust for DASS anxiety and DASS stress levels in addition to age, gender, and education levels as covariates. This model was also significant, $F_{6,92}$=4.05, P=.001 and explained about 21% of the variance in cybersickness severity (R^2=0.21, adjusted R^2=0.16). The only predictor emerging as significant was DASS depression (β=1.15; 95% CI 0.29 to 2.01; P=.009), and DASS anxiety reached marginal significance (β=1.2; 95% CI -0.07 to 2.47; P=.06). Cybersickness severity predictions showed a similar trend as for the PHQ-9 scores - Figure 8 (B).

Figure 8. Predicted levels of cybersickness severity (and 95% CIs) on the Simulator Sickness Questionnaire along depressive symptoms severity measured via (A) the Patient Health Questionnaire and (B) the Depression Anxiety Stress Scales.



Notes. Depressive symptom severity scores here are limited to the range of data that was available and do not cover the whole range of the PHQ-9 or DASS-D scores.
Model B adjusts for self-reported baseline differences in anxiety and stress levels based on the DASS questionnaire.

Discussion

The current study reports on the acceptability, usability and cybersickness levels of a novel VR

environment designed for the assessment of depressive symptoms. Following a circa 30 min engagement, 50 healthy controls and 50 individuals with moderate self-reported depressive symptom severity completed a self-report battery on the three constructs of interest.

Acceptability and usability

VR technology was found acceptable by the majority of participants; both for the purpose of mental health screening as well as treatment. For diagnostics, a clear preference emerged for VR to be used as a support tool by a healthcare professional, as opposed to a stand-alone solution – which is in line with previous findings concerning digital solutions [39]. The high level of acceptance users had shown towards this technology coupled with its advantages lays a solid ground for investigations into how VR could be utilized in a mental health context. VR being a premise for real-time data collection via direct observation of behavior contributes to higher ecological validity and allows for the consideration of a wider set of variables when assessing mental states [6,7]. While the potential is great, reviews of real-life studies (e.g. [40]), editorials (e.g. [41]) as well as commentaries (e.g. [42]), highlight a need for further scientific investigation, to better inform how VR is best implemented within clinical practice. For these purposes knowledge on user perspectives is key.

The specific pilot system tested here was designed for the purpose to differentiate between healthy controls and those with depressive symptoms (results to be published). The system itself was rated acceptable and usable by most participants, but the need for improvement was apparent; specifically, when it came to the users' confidence levels. These results call for more attention from the scientific community, as depression has been understudied compared to other psychiatric conditions when it comes to the utilization of VR [42,43], while participants here indicated support for the technology both in theory and regarding this specific pilot system. Identifying novel approaches with user support is particularly important for groups that might find treatment adherence and lack of motivation to engage challenging, such as those with depression [44].

Cybersickness

Finally, results of the study show significant differences in self-reported cybersickness levels between healthy controls and those with depressive symptoms – the latter reporting considerably higher levels. This difference brings forward many questions, some relating to the self-report instrument used, as well as its administration; these will be discussed under the limitations more thoroughly. Beyond the critical appraisal of the instrument, this difference also urges for a consideration of how cybersickness might relate to depression. It is important to consider that our study lacked a pre-post comparison, which creates challenges when attempting to disentangle cybersickness symptoms evoked by VR exposure from potential baseline differences stemming from somatic symptoms related to depression. While "the language used in the medical literature to describe somatic symptoms in depression is both confusing and contradictory" [45] one can pinpoint constructs related to depression, that may also be captured by cybersickness scales, which may cause conflation when assessing cybersickness levels. These include, but might not be limited to headache, fatigue, nausea, dizziness and gastrointestinal disturbances (e.g. [23,45]).

Nonetheless, it is important to consider that the elevated cybersickness in the group with depressive symptoms could still be, at least in part, a response to the VR environment itself. If this is true, using VR with the right study design may serve as a potential tool for further exploring the sensory abnormalities in depression. Investigating the underlying mechanisms of cybersickness in depression may also provide novel insights into the somatic manifestations of the condition and could help further define depression as well as refine interventions for mental health. Additionally, the need to consider cybersickness severity becomes even more apparent in light of the results showing its

association with acceptability and usability ratings, which might transfer into the real-life uptake and use of VR technology.

Limitations and future directions

The study is not without its limitations. First, one has to consider the characteristics of participants, as the sample consisted of young adults, the majority being women, and highly educated which limits generalizability [46]. However, further reports are needed to understand how exactly such sample characteristics - sociodemographic variables as well as depressive symptoms - influence user attitudes to digital technologies and engagement in a mental health context [47,48]. Second, the categorization between the 'healthy control' and the 'depressed' groups was based solely on self-reported symptoms at the time of the study (using the PHQ-9) as opposed to the presence or absence of a clinical diagnosis. Additionally, the majority of individuals labelled as 'depressed' only reported moderate depression severity, which questions whether the results are applicable to those with more severe symptoms. Third, the case and control groups differed in self-reported stress and anxiety levels, which highlights that between-group differences cannot unequivocally be attributed to difference in depressive symptom severity. Fourth, the measurement of cybersickness levels leaves room for improvement on multiple fronts. The instrument (SSQ) has repeatedly been criticized for reasons among others being the low power to differentiate from anxiety, and the questionable assumption of having zero 'symptoms' at baseline [24–27]. The difference in severity between the healthy controls and those with depressive symptoms cannot unequivocally be attributed to the VR exposure, as no baseline levels were recorded. It is possible, that at least some degree of the difference originates from somatic symptoms of depression, which the questionnaire most likely does not adequately differentiate from cybersickness symptoms. In the future, assessment of related symptoms necessitates pre-post comparisons. Fifth, when creating items to assess constructs of the TFA, one item was taken from the SUS questionnaire (Item 9 – Confidence), while all other items were specialized to this study. This item showed response patterns different from all other items within the TFA, leading to questionable reliability. Sixth, the results of the ordinal regression models should be interpreted with caution, especially in cases when the assumption of proportional odds was not fulfilled. This suggests that the relationship between predictors and the outcome variable may differ across categories and thus future studies could expand on these findings by exploring more flexible modeling approaches. Last, one must mention that consideration of clinician perspectives is equally important to the patient experience. As such, future studies could consider involving other user groups, and extend the consideration to implementation, feasibility and cost-effectiveness of VR-based solutions to enable a more comprehensive evaluation.

Conclusion

There is a clear need to enhance mental health diagnosis, potentially by incorporating a broader range of variables. This study explored a novel approach to increase the ecological validity of depression assessments and found that, while VR technologies are generally acceptable as a supplementary tool, they are not seen as a replacement for routine mental health evaluations. The study, which tested the initial version of a novel VR system, revealed that majority of participants found the acceptability and usability satisfactory, despite experiencing considerable levels of cybersickness. Notably, the results highlighted previously unreported differences in cybersickness severity between individuals with depressive symptoms and healthy controls, warranting replication and further investigation. While the call for enhanced reliability in mental health assessments is well-founded, and novel approaches are under investigation for their efficacy and accuracy, greater emphasis must be placed on evaluating acceptability and usability. Ensuring a safe, acceptable and user-friendly approach is essential for the successful implementation of these technologies beyond the testing phase.

Conflict of interest

None to be declared.

Acknowledgement

The study was carried out within the EXPERIENCE project, which is funded by the European Commission H2020 Framework Program, Grant No. 101017727.

References

1. Collaborators G 2019 MD. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry*. 2022;9(2):137-150. doi:10.1016/s2215-0366(21)00395-3

2. Friedrich MJ. Depression Is the Leading Cause of Disability Around the World. *JAMA*. 2017;317(15):1517-1517. doi:10.1001/jama.2017.3826

3. Proudman D, Greenberg P, Nellesen D. The Growing Burden of Major Depressive Disorders (MDD): Implications for Researchers and Policy Makers. *PharmacoEconomics*. 2021;39(6):619-625. doi:10.1007/s40273-021-01040-7

4. Lieblich SM, Castle DJ, Pantelis C, Hopwood M, Young AH, Everall IP. High heterogeneity and low reliability in the diagnosis of major depression will impair the development of new drugs. *BJPsych Open*. 2015;1(2):e5-e7. doi:10.1192/bjpo.bp.115.000786

5. Liu X, Jiang K. Why is Diagnosing MDD Challenging? *Shanghai Arch Psychiatry*. 2016;28(6):343-345. doi:10.11919/j.issn.1002-0829.216073

6. Bell IH, Nicholas J, Alvarez-Jimenez M, Thompson A, Valmaggia L. Virtual reality as a clinical tool in mental health research and practice. *Dialogues Clin Neurosci*. 2020;22(2):169-177. doi:10.31887/dcns.2020.22.2/lvalmaggia

7. Freeman D, Reeve S, Robinson A, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol Med*. 2017;47(14):2393-2400. doi:10.1017/s003329171700040x

8. Mao K, Wu Y, Chen J. A systematic review on automated clinical depression diagnosis. *npj Ment Heal Res*. 2023;2(1):20. doi:10.1038/s44184-023-00040-z

9. Bilello JA. Seeking an objective diagnosis of depression. *Biomark Med*. 2016;10(8):861-875. doi:10.2217/bmm-2016-0076

10. Abd-Alrazaq A, AlSaad R, Shuweihdi F, Ahmed A, Aziz S, Sheikh J. Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. *npj Digit Med*. 2023;6(1):84. doi:10.1038/s41746-023-00828-5

11. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: A systematic review and meta-analysis. *J Affect Disord*. 2014;152:45-51. doi:10.1016/j.jad.2013.10.002

12. DiMatteo MR, Lepper HS, Croghan TW. Depression Is a Risk Factor for Noncompliance With Medical Treatment: Meta-analysis of the Effects of Anxiety and Depression on Patient Adherence. *Arch Intern Med*. 2000;160(14):2101-2107. doi:10.1001/archinte.160.14.2101

13. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of

reviews and development of a theoretical framework. *BMC Heal Serv Res*. 2017;17(1):88. doi:10.1186/s12913-017-2031-8

14. Sauer J, Sonderegger A, Schmutz S. Usability, user experience and accessibility: towards an integrative model. *Ergonomics*. 2020;63(10):1207-1220. doi:10.1080/00140139.2020.1774080

15. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a Brief Depression Severity Measure. *JGIM*. 2001;16(9):606–613. doi:10.1046/j.1525-1497.2001.016009606.x

16. Lovibond PF, Lovibond SH. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther*. 1995;33(3):335-343. doi:10.1016/0005-7967(94)00075-u

17. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*. 1988;6(54):1063-1070.

18. Kashdan TB, Gallagher MW, Silvia PJ, et al. The curiosity and exploration inventory-II: Development, factor structure, and psychometrics. *J Res Pers*. 2009;43(6):987-998. doi:10.1016/j.jrp.2009.04.011

19. Sekhon M, Cartwright M, Francis JJ. Development of a theory-informed questionnaire to assess the acceptability of healthcare interventions. *BMC Heal Serv Res*. 2022;22(1):279. doi:10.1186/s12913-022-07577-3

20. Brooke J. SUS: A quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, eds. *Usability Evaluation in Industry*. ; 1996:189-194.

21. Kortum P, Peres SC. Evaluation of Home Health Care Devices: Remote Usability Assessment. *JMIR Hum Factors*. 2015;2(1):e10. doi:10.2196/humanfactors.4570

22. Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*. 2009;4:114-123.

23. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*. 1993;3(3):202-220. doi:10.1207/s15327108ijap0303_3

24. Bimberg P, Weissker T, Kulik A. On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research. *2020 IEEE Conf Virtual Real 3D User Interfaces Abstr Work (VRW)*. 2020;00:464-467. doi:10.1109/vrw50115.2020.00098

25. Bouchard S, Berthiaume M, Robillard G, et al. Arguing in Favor of Revising the Simulator Sickness Questionnaire Factor Structure When Assessing Side Effects Induced by Immersions in Virtual Reality. *Front Psychiatry*. 2021;12:739742. doi:10.3389/fpsyt.2021.739742

26. Brown P, Spronck P, Powell W. The simulator sickness questionnaire, and the erroneous zero baseline assumption. *Front Virtual Real*. 2022;3:945800. doi:10.3389/frvir.2022.945800

27. Sevinc V, Berkman MI. Psychometric evaluation of Simulator Sickness Questionnaire and its

variants as a measure of cybersickness in consumer virtual environments. *Appl Ergon*. 2020;82:102958. doi:10.1016/j.apergo.2019.102958

28. Posit_team. *RStudio: Integrated Development Environment for R*. PBC; 2024. http://www.posit.co/

29. Wickham H, François R, Henry L, Müller K, Vaughan D. *Dplyr: A Grammar of Data Manipulation*.; 2023. https://dplyr.tidyverse.org

30. Adler D, Kelly ST, Elliott T, Adamson J. *Vioplot: Violin Plot*.; 2024. https://github.com/TomKellyGenetics/vioplot

31. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016. https://ggplot2.tidyverse.org

32. Revelle W. *Psych: Procedures for Psychological, Psychometric, and Personality Research*.; 2024. https://CRAN.R-project.org/package=psych

33. Kassambara A. *Rstatix: Pipe-Friendly Framework for Basic Statistical Tests*.; 2023. https://rpkgs.datanovia.com/rstatix/

34. Lüdecke D. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*. 2018;3(26):772. doi:10.21105/joss.00772

35. Christensen RHB. *Ordinal - Regression Models for Ordinal Data*.; 2023. https://CRAN.R-project.org/package=ordinal

36. Yee TW. The VGAM Package for Categorical Data Analysis. *Journal of Statistical Software*. 2010;32(10):1-34. doi:10.18637/jss.v032.i10

37. Carli V, Gentili C. EXPERIENCE: Is virtual reality suitable to identify differences between depressed and healthy individuals? 2022. https://doi.org/10.1186/isrctn16396369

38. Saredakis D, Szpak A, Birckhead B, Keage HAD, Rizzo A, Loetscher T. Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Front Hum Neurosci*. 2020;14:96. doi:10.3389/fnhum.2020.00096

39. Chan AHY, Honey MLL. User perceptions of mobile digital apps for mental health: Acceptability and usability - An integrative review. *J Psychiatr Ment Heal Nurs*. 2022;29(1):147-168. doi:10.1111/jpm.12744

40. Geraets CNW, Wallinius M, Sygel K. Use of Virtual Reality in Psychiatric Diagnostic Assessments: A Systematic Review. *Front Psychiatry*. 2022;13:828410. doi:10.3389/fpsyt.2022.828410

41. Riva G, Serino S. Virtual Reality in the Assessment, Understanding and Treatment of Mental Health Disorders. *J Clin Med*. 2020;9(11):3434. doi:10.3390/jcm9113434

42. Selaskowski B, Wiebe A, Kannen K, et al. Clinical adoption of virtual reality in mental health is

challenged by lack of high-quality research. *npj Ment Heal Res*. 2024;3(1):24. doi:10.1038/s44184-024-00069-8

43. Cieślika B, Mazurekb J, Rutkowskic S, Kiperd P, Turollad A, Szczepańska-Gierachae J. Virtual reality in psychiatric disorders: A systematic review of reviews. *Complementary Therapies in Medicine*. 2020;52:102480. doi:10.1016/j.ctim.2020.102480

44. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Évid Based Ment Heal*. 2018;21(3):116. doi:10.1136/eb-2018-102891

45. Tylee A, Gandhi P. The Importance of Somatic Symptoms in Depression in Primary Care. *Primary care companion to the Journal of clinical psychiatry*. 2005;7(4):167-176. doi:10.4088/pcc.v07n0405

46. Borghouts J, Eikey E, Mark G, et al. Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review. *Journal of medical Internet research*. 2021;23(3):e24387. doi:10.2196/24387

47. Chudy-Onwugaje K, Abutaleb A, Buchwald A, et al. Age Modifies the Association Between Depressive Symptoms and Adherence to Self-Testing With Telemedicine in Patients With Inflammatory Bowel Disease. *Inflamm Bowel Dis*. 2018;24(12):2648-2654. doi:10.1093/ibd/izy194

48. Abouzeid N, Lal S. The role of sociodemographic factors on the acceptability of digital mental health care: A scoping review protocol. *PLOS ONE*. 2024;19(4):e0301886. doi:10.1371/journal.pone.0301886

Appendicies

Supplementary Material 1.

Questionnaire on acceptability based on the Theoretical Framework of Acceptability [13,19].
The questionnaire was administered in Italian due to the site of recruitment being the University if Padua, Italy.

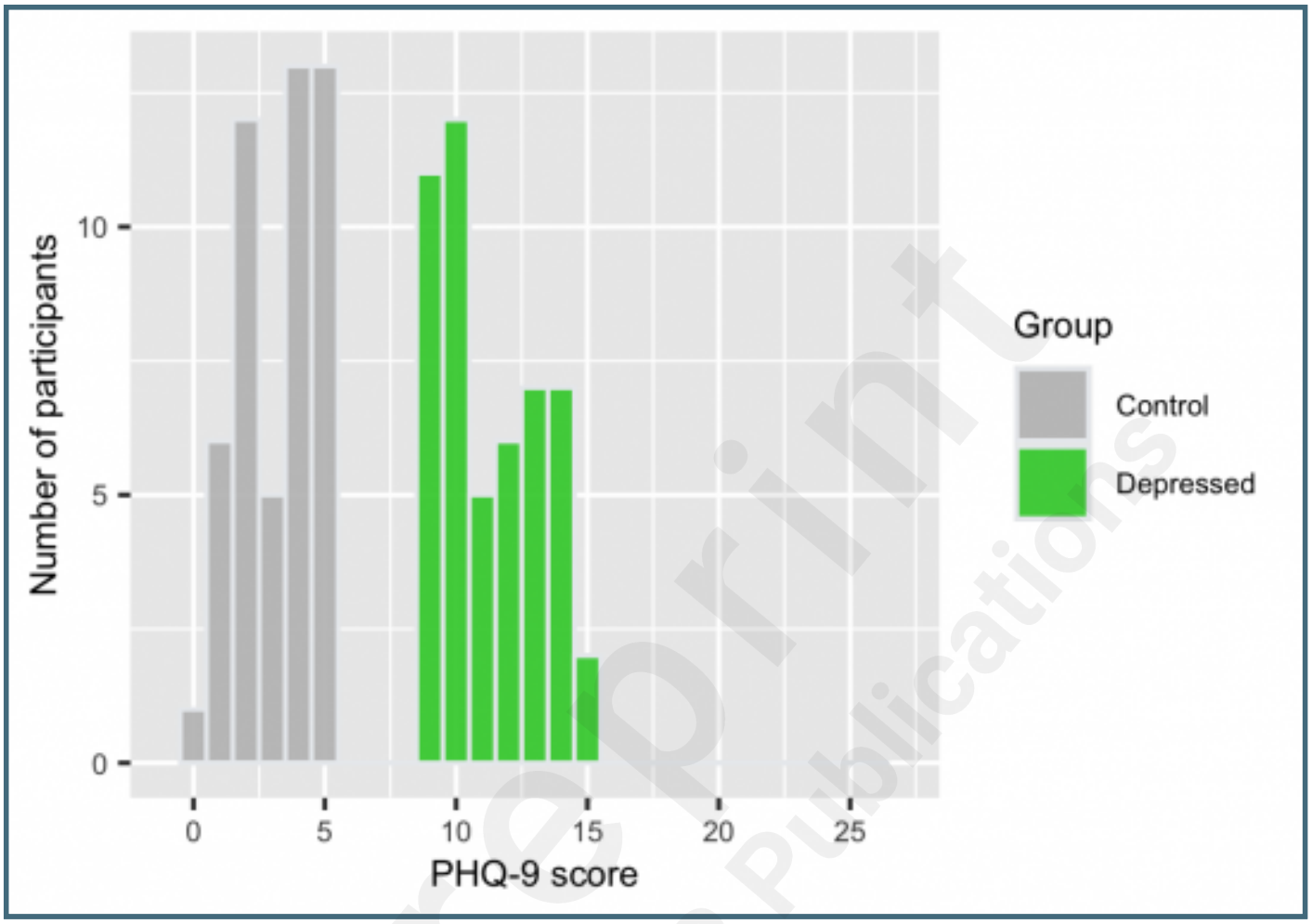
Supplementary Material 2.

Script used for the analysis and visualization, including its output.

Supplementary Files

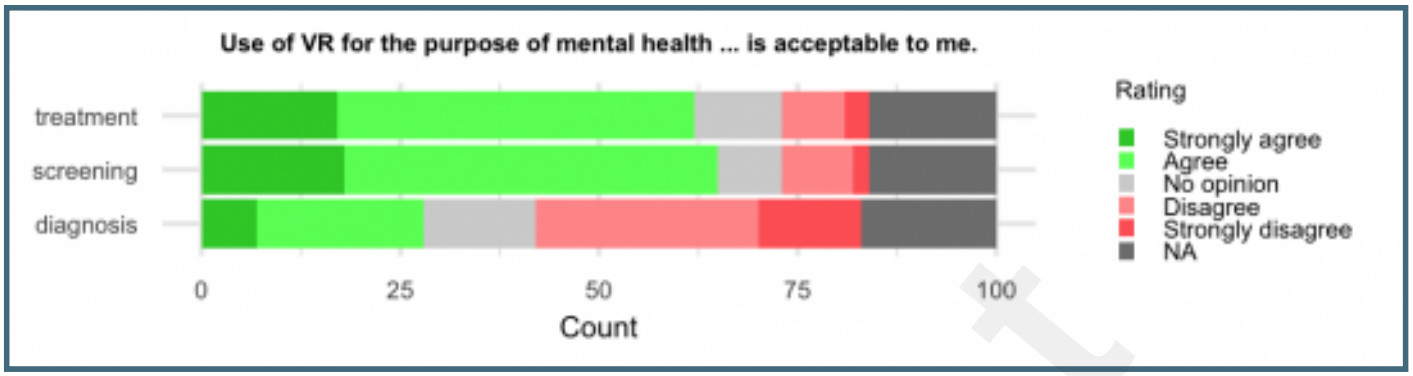
Figures

Distribution of participants' scores on the Patient Health Questionnaire (PHQ-9) assessing current depressive symptom severity.

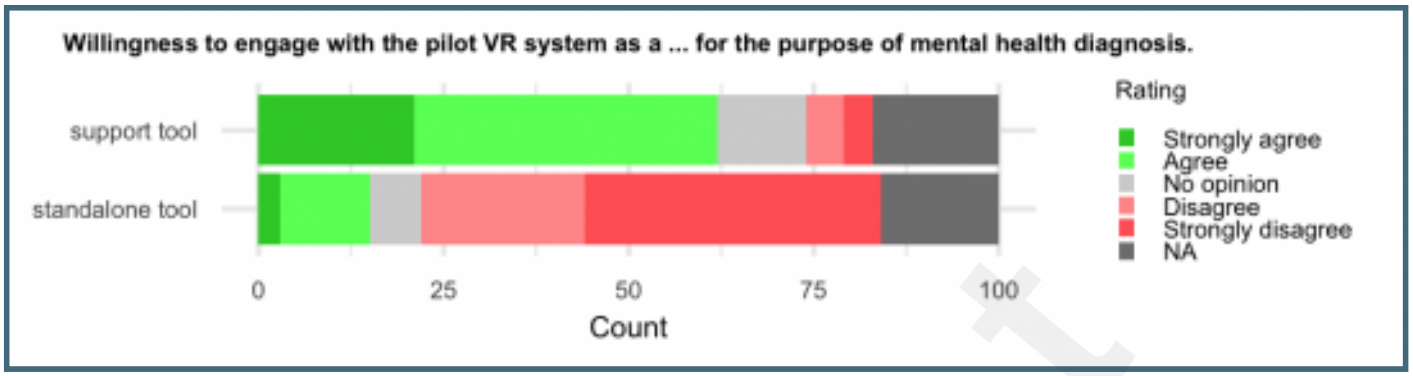


The visual illustration of the virtual reality environment designed for the assessment of cognition and behavior hypothesized to be related to depression.

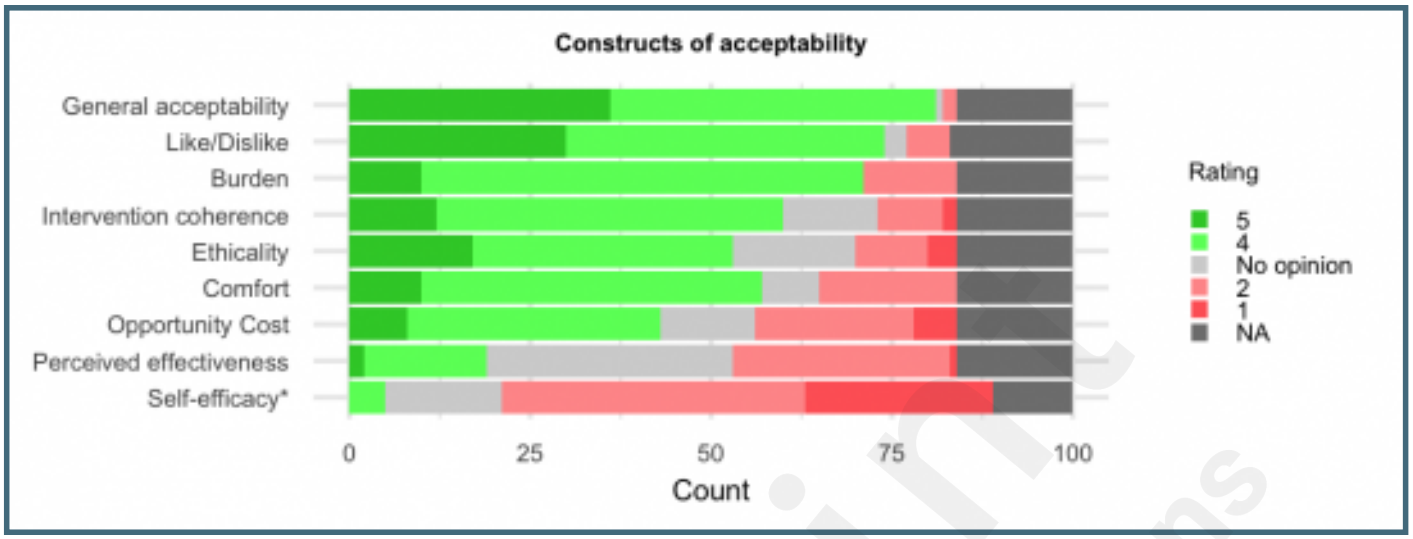
Distribution of scores on the acceptance of Virtual Reality technology for the purpose of mental health screening, diagnosis and treatment.



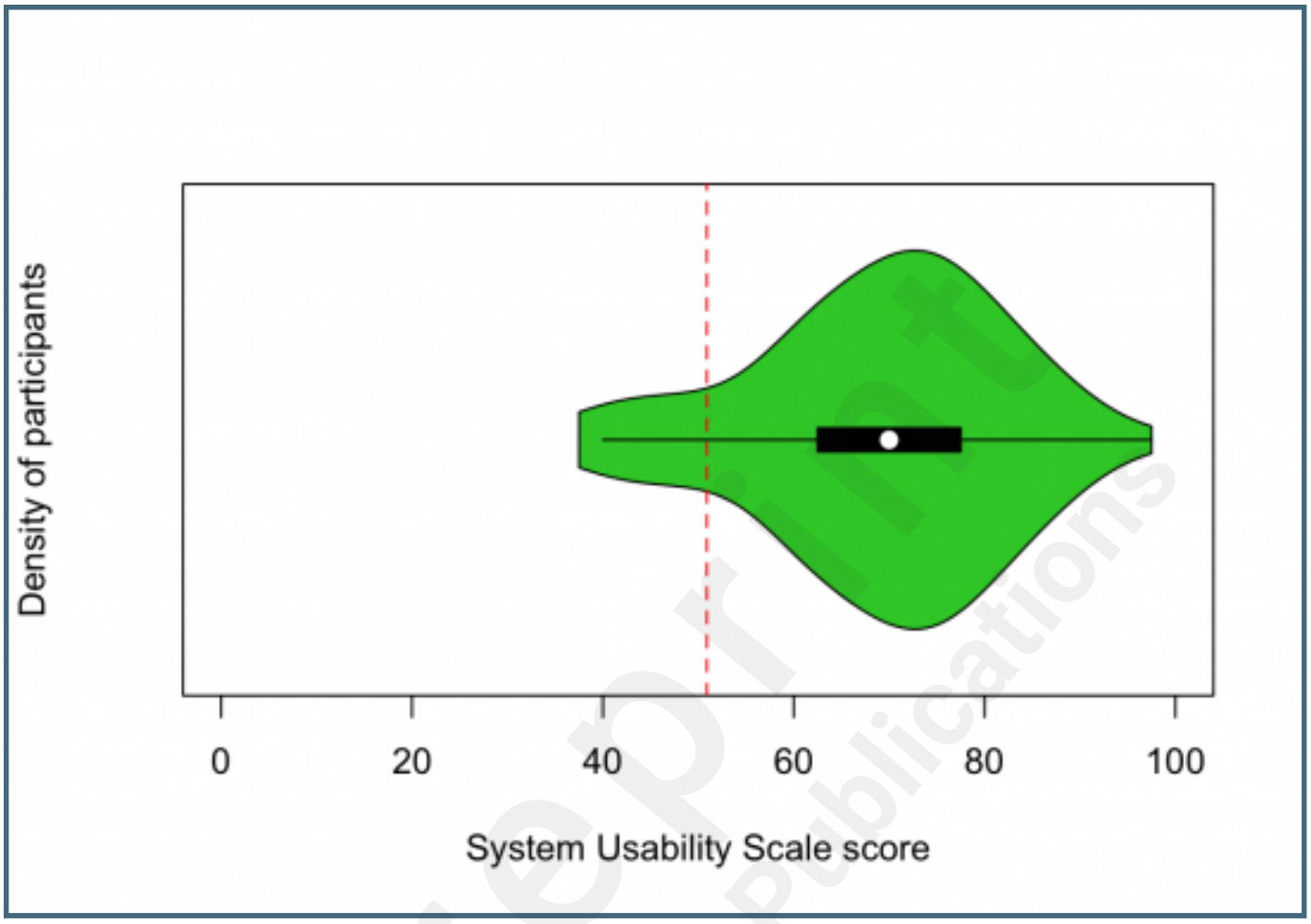
Willingness to engage with the pilot VR system for the purpose to receive a diagnosis with (support tool) and without (standalone tool) input from a mental health professional.



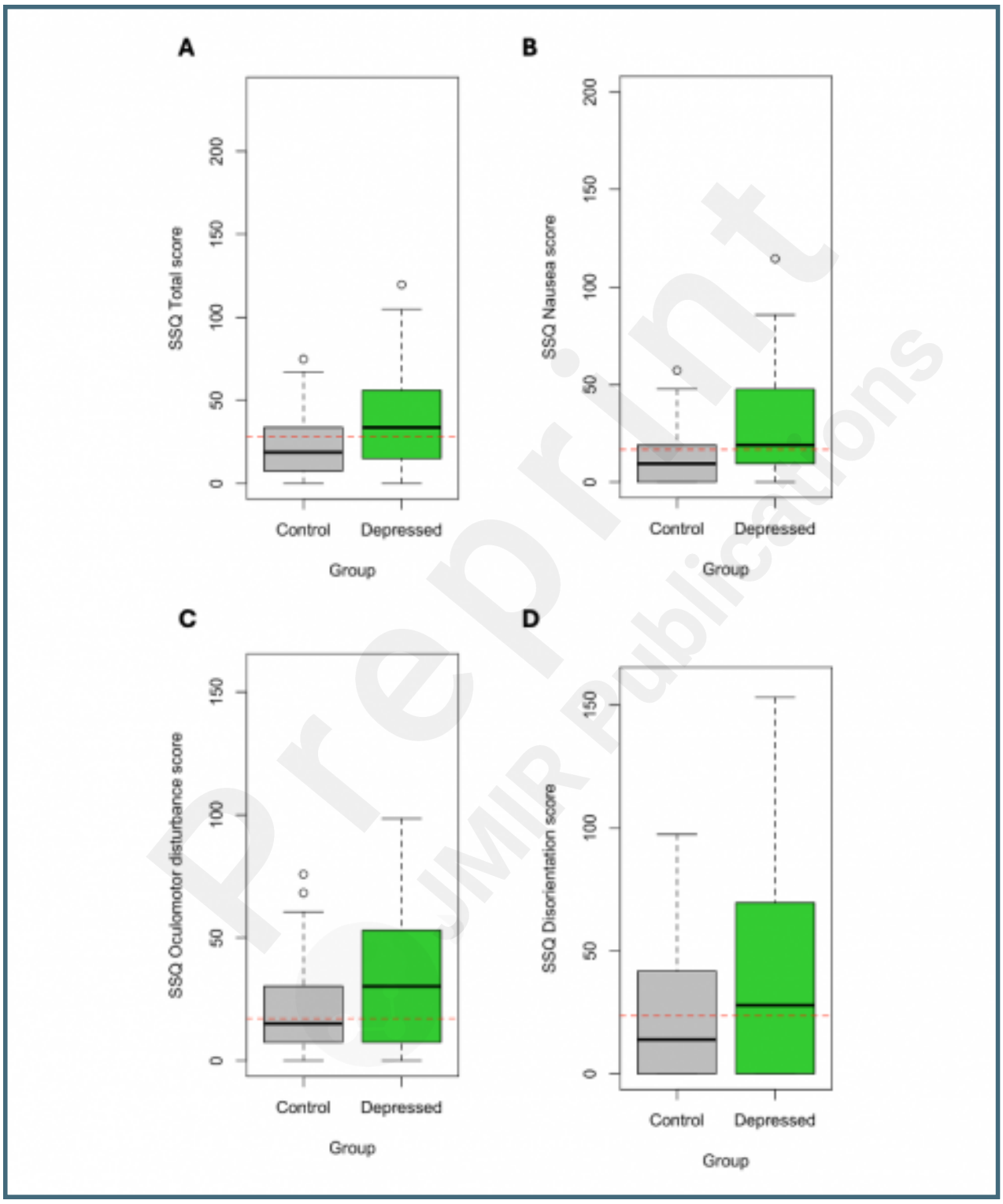
Self-reported acceptability of the pilot VR system along the constructs proposed within the Theoretical Framework of Acceptability.



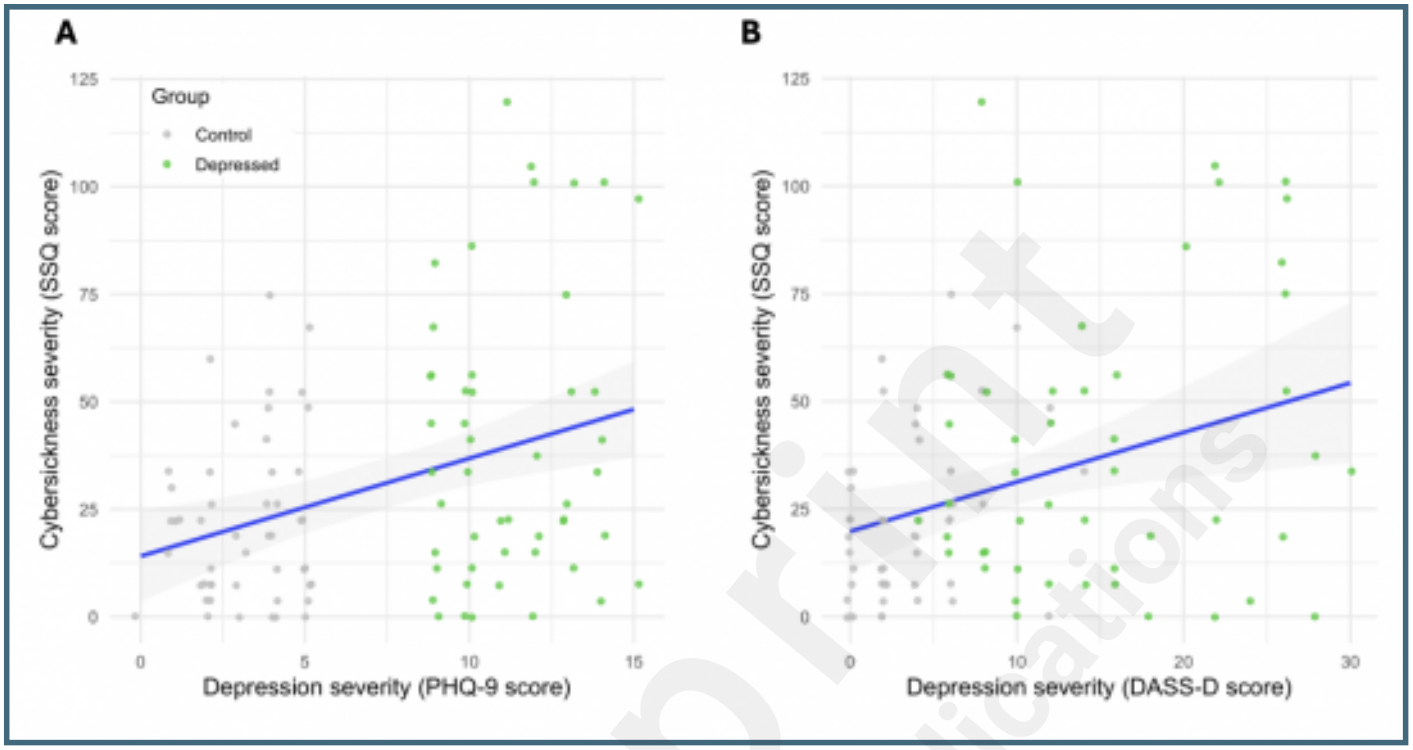
Density plot of usability ratings from all participants.



Boxplot of self-reported cybersickness severity scores of participants per group on the Simulator Sickness Questionnaire (A), and its three subscales of Nausea (B), Oculomotor disturbances (C) and Disorientation (D).



Predicted levels of cybersickness severity (and 95% CIs) on the Simulator Sickness Questionnaire along depressive symptoms severity measured via (A) the Patient Health Questionnaire and (B) the Depression Anxiety Stress Scales.



Multimedia Appendixes

Questionnaire on acceptability based on the Theoretical Framework of Acceptability.
URL: http://asset.jmir.pub/assets/d498721b690b72bf3610fcfe3ea0d88f.docx

Script used for the analysis and visualization, including its output.
URL: http://asset.jmir.pub/assets/8d34fe7a4472d13db3f2357738d1d948.pdf