JMIR XR and Spatial Computing (JMXR)

Volume 2 (2025) ISSN 2818-3045 Editor in Chief: Lars Riedemann, MD

Contents

Reviews

Original Papers

Immersive Virtual Reality for Health Promotion and Primary Prevention in Psychology: Scoping Review

Solenne Bonneterre[1], PhD; Oulmann Zerhouni[2], Prof Dr; Marilisa Boffo[3], PhD

[1]Laboratoire Parisien de Psychologie Sociale, Département de Psychologie, Université Paris Nanterre, Nanterre, France

[2]Centre de Recherche sur les Fonctionnements et Dysfonctionnements Psychologiques, UR 7475, Université de Rouen Normandie, Rouen, France

[3]Department of Psychology, Education, and Child Studies, Erasmus School of Social and Behavioral Sciences, Erasmus University Rotterdam, Rotterdam, Netherlands

Corresponding Author:
Solenne Bonneterre, PhD
Laboratoire Parisien de Psychologie Sociale, Département de Psychologie, Université Paris Nanterre, Nanterre, France

Abstract

Background: Virtual reality (VR) has emerged as a promising tool in health promotion and prevention psychology. Its ability to create immersive, engaging, and standardized environments offers unique opportunities for interventions and assessments. However, the scope of VR applications in this field remains unclear.

Objective: This scoping review aims to identify and map the applications of VR in health promotion and prevention psychology, focusing on its uses, outcomes, and challenges.

Methods: A systematic search was conducted across 3 electronic databases (PubMed, PsycINFO, and Scopus) for studies published between 2010 and 2024. Eligibility criteria included empirical studies using immersive VR for health promotion and prevention, while studies using nonimmersive VR, lacking health-related applications, or focusing on clinical interventions were excluded. The review followed PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews) guidelines, and 4295 records were initially identified, with 51 studies included after screening. Data were synthesized qualitatively to identify key applications, limitations, and emerging trends.

Results: VR was primarily used in three areas: (1) delivering interventions (eg, pilot testing, skills training), (2) exploring fundamental research questions, and (3) assessing outcomes such as behavioral or psychological responses. Although VR demonstrated potential for enhancing user engagement and replicating ecological scenarios, its effectiveness compared to nonimmersive methods varied. Most studies were pilot or feasibility studies with small, nonrepresentative samples, short follow-up periods, and limited methodological standardization.

Conclusions: VR offers a versatile and promising tool for health promotion and prevention but its applications are still in the early stages. The evidence is limited by methodological weaknesses and variability in outcomes. Future research should prioritize replication, longitudinal designs, and standardized methodologies to strengthen the evidence base and expand the applicability of VR interventions.

(JMIR XR Spatial Comput 2025;2:e49923) doi:10.2196/49923

KEYWORDS

virtual reality; health psychology; prevention psychology; health promotion

Introduction

Background

Health and prevention psychology aims to address health-related issues to either prevent individuals from starting or continuing an unhealthy behavior (ie, primary prevention), help them to detect or reduce illness in early stages (ie, secondary prevention), or support individuals in their journey against consequences of heavier injuries or diseases (ie, tertiary prevention, [1]). Although secondary and tertiary prevention are more individual-based depending on the illness or signs or symptoms individuals need to learn to cope with, primary prevention is broader and aimed at a larger audience. Therefore, primary or universal prevention is designed to prevent individuals from the general population from getting injured or sick and aims to enable people to live a sustainable and healthy lifestyle [2,3].

In this sense, health promotion campaigns have started to integrate technological innovations such as virtual reality (VR). We refer to VR as a type of human-computer interface immersing users into a computer-generated 3D virtual environment (VE) they can interact with in a naturalistic fashion, usually via an avatar (ie, representation of the user in the VE [4]). More pragmatically, we labeled as VR any type of device that has the ability to sensorily detach the user from the outside

world (at least sight, but also sounds, smell, and touch in some cases). This includes the use of a cave automatic VE (users are surrounded by walls displaying the VE) or a head-mounted display (HMD), which blocks the user's field of view outside of the VE and from which the user cannot turn away by simply looking away (ie, computer screens or 360° videos will not be considered VR in this definition).

The main aim of VR is to recreate a realistic, ecological context and experience while keeping some degree of experimental control over it [5-7]. Systematic reviews have reported promising results from VR-based interventions in other disciplines (eg, clinical psychology [8] and social psychology [9,10]). However, to our knowledge, there has been no review of the use of VR technologies for primary health promotion and prevention. Therefore, instead of focusing on specific research questions related to a topic, outcome, or population, the goal of this review was to map the current state of the art of the use of VR in such areas and identify gaps and future directions.

Rationale

Virtual Reality: Operating Principles

The VR literature highlights 2 essential concepts, immersion and presence, both of which are critical to the user's experience in VEs [7,11]. Immersion refers to the technological ability of a VR system to fully engage the user by replacing real-world sensory inputs with virtual stimuli. The more immersive the device, the less interface there is between the user and the virtual world. High immersion includes naturalistic interactions, such as the use of body suits to track movement, which increases the sense of realism [7]. Immersive systems create a sense that the virtual world is an actual experience rather than a mediated one. However, presence depends on the user's psychological response to the VE. It is the subjective feeling of "being there" in the virtual world, interacting with it as if it were real [12]. This sense of presence increases engagement and leads to more vivid, memorable experiences [13]. Notably, presence can be felt in both immersive and nonimmersive media, such as movies or books, as it is influenced by individual factors and not just the technological features of the medium [14].

Although immersion and presence are often related, they are not the same. Higher levels of immersion tend to enhance feelings of presence, but immersion is not a necessary condition for presence [15]. Thus, immersion can be viewed as a moderator that enhances presence but does not guarantee it [16].

Why Use VR in Health Promotion and Prevention Psychology?

VR technology has emerged as a promising tool in health promotion and prevention psychology, allowing for immersive experiences that can enhance user engagement and motivation [17,18]. VR enables researchers to create safe, ecological, and standardized VEs, where health promotion interventions can be effectively delivered and evaluated. VR presents key advantages as a tool for research and intervention in health promotion and primary prevention [7].

First, VR can be combined with devices aimed at mimicking more natural movements (eg, the use of handheld controllers or haptic devices instead of a mouse and keyboard) and can encompass the integration of full-body motor and haptic feedback when using a bodysuit. This freedom and wholeness of movement can help enhance learning through direct practice, visualization, and ultimately embodied cognition (ie, cognition linked to the body [7,19]). Hence, VR can be a relevant tool to create interventions aimed at learning health-related behaviors that require practicing skills (eg, detecting testicular disorders [20]).

Second, due to its ability to elicit embodiment, VR is well suited to elicit and enhance perspective-taking and empathy [7,21]. For example, embodying an obese avatar could enhance taking the perspective of being overweight, leading to a more effective learning of the consequences of obesity and, in turn, a greater intention to take care of individual health (ie, reduce the attitude-intention-behavior gap). Through the feeling of presence, individuals can visualize themselves in a specific situation, hence allowing a deeper sense of self-reflection [20], potentially leading to more persistent changes in behavior. VR can recreate ecological situations and environments in which users can embody an avatar and act in the virtual world as if it were real, through the feeling of presence [12].

Objective

Our goal was to identify and map how VR has been used in the field of health promotion and primary prevention. In this scoping review, we addressed three broad research questions:

1. What are the uses of VR technology in primary prevention and health promotion (ie, an overview of the goals and research questions addressed through the use of VR)?
2. What do we know so far about the effects of using VR in these fields (ie, a summary of the results)?
3. What are the challenges and limitations, if any, encountered so far?

Based on the findings of the scoping review, we drafted a list of recommendations and perspectives for the use of VR in health promotion and primary prevention.

Methods

Protocol and Registration

The scoping review protocol was drafted according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews) checklist [22,23]. We also conducted a synthesis without meta-analysis [24] (Checklist 1).

Eligibility Criteria

We included any peer-reviewed and published empirical article, written in English, that described a study conducted on human subjects deploying any kind of immersive VR device (eg, HMD, cave automatic VE), including 360° videos when used in a VR setup, focusing on any research question in the field of health promotion or primary prevention, from January 1, 2010, to September 16, 2024. We chose to limit the search to the last 14 years in order to generate a recent state-of-the-art overview of the field. We excluded studies conducted on nonhumans or focused on secondary or tertiary prevention interventions, such

as psychotherapeutic treatments (eg, VR exposure therapy) and medical interventions (eg, rehabilitation), or specialized educational programs unrelated to prevention (eg, skills improvement for health practitioners). Pilot studies were not excluded from this review because of their critical role in assessing the feasibility and acceptability of interventions that may inform future primary, secondary, or tertiary prevention efforts. We excluded studies using the term "virtual reality" that described computer-based VEs involving a virtual world (eg, Second Life) or computer-related or motion-sensing devices (eg, Kinect, joystick) when they were associated with a nonimmersive VR setup (eg, non-VR video or serious game). We also used the population-concept-context framework to define our inclusion criteria. The population includes adolescents, young adults, and specific populations at risk for health issues (eg, individuals with anxiety or those at risk for substance use). The concept focuses on the application of VR technology to promote health behaviors, enhance knowledge, and improve emotional well-being. The context refers to contextual factors including the environments where VR interventions are delivered, such as schools, community centers, or health care facilities.

Information Sources and Search Process

We searched 3 databases from January 1, 2010, until September 16, 2024 (PubMed and PsycINFO). For each database, we combined 2 sets of keywords; the first set focused on health promotion and prevention psychology. For PubMed, the search strings were ("health prevention" OR "health promotion" OR "health risk communication" OR "health communication" OR "preventive psychology" OR "behavior change" OR "attitude change") AND ("virtual reality" OR "immersive virtual reality" OR "immersive virtual environment"). For PsycINFO, the search strings were ("health prevention" OR "health promotion" OR "health risk communication" OR "health communication" OR "preventive psychology" OR "behavior change" OR "attitude change") AND ("virtual reality" OR "immersive virtual reality" OR "immersive virtual environment").

Selection of Sources of Evidence

Studies that did not employ VR technology, were not peer-reviewed, were reviews or meta-analyses, or lacked empirical data were excluded from the review. The screening process was conducted in 2 stages to enhance the rigor of the selection. In the first stage, titles and abstracts of the identified studies were reviewed to determine their relevance based on the inclusion criteria. This initial screening allowed the authors to eliminate studies that were clearly outside the scope of the review. In the second stage, full-text articles of the remaining studies were assessed to confirm their eligibility for inclusion. The extraction process was conducted independently by multiple reviewers to enhance reliability and minimize bias. Any discrepancies in data extraction were resolved through discussion and consensus among the reviewers. This meticulous approach to data extraction allowed the authors to synthesize findings across studies effectively and draw meaningful conclusions regarding the efficacy and feasibility of VR interventions in health promotion and primary prevention.

Data Charting Process

The data charting process involved collecting information on study characteristics, intervention details, measured outcomes, user experience, type of materials, and sample characteristics (see Multimedia Appendix 1). Of note, approximately 63% of the studies included in the review were categorized as pilot or feasibility studies. We also recorded the type of VR technology used (eg, immersive headsets, desktop VR), the duration of the intervention, and the focus of the VR content (eg, health education, behavior change). On average, participants spent approximately 12.8 (SD 11.1) minutes using VR. We focused on health-related outcomes such as knowledge acquisition, behavioral intentions, and psychological well-being. User experience was assessed through qualitative data that provided insights into participants' enjoyment, ease of use, and perceived effectiveness of the VR interventions. Many studies found that participants found the VR experience both enjoyable and engaging, which in turn led to higher participation rates compared to non-VR interventions.

Data Items

Primary variables included study characteristics such as authorship, year of publication, study design, and sample size, which provided context for the research findings. Participant demographics, including age, gender, and health status, were also collected to understand the populations included in the studies. Intervention details were documented, focusing on the type of VR technology used, the duration of the intervention, and the specific health issues addressed. Measured outcomes were categorized into primary outcomes, such as knowledge acquisition and behavioral intentions, and secondary outcomes, including user engagement and satisfaction. User experience data were collected to assess participants' enjoyment, ease of use, and any challenges encountered during the VR interventions. In addition, limitations of the studies were noted, including issues such as small sample sizes and methodological limitations, which are critical for contextualizing the findings. It is important to note that while immersion and presence are key concepts in understanding the effectiveness of VR, these variables were not measured consistently across studies, which may affect the interpretation of results. The data elements collected were intended to provide a structured review of the existing literature, as well as identify trends, gaps, and implications for future research in the field of VR-based health interventions.

Critical Appraisal of Individual Sources of Evidence

We found that approximately 37% of the included studies were pilot or feasibility studies. These studies primarily focused on evaluating the usability and acceptability of VR interventions, which are critical for assessing the feasibility of larger-scale research. Although pilot studies provide valuable insights into user experiences and preliminary results, their small sample sizes and limited generalizability limit the ability to draw firm conclusions about the effectiveness of VR-based interventions. Mixed results have been found when comparing VR interventions to traditional methods, suggesting that VR does not always offer a clear advantage in achieving health outcomes. Key variables such as immersion and presence, which are critical

to understanding how VR might influence health behaviors, have not been systematically evaluated. We found a lack of focus on larger, more diverse samples and aim to replicate existing studies to strengthen the evidence supporting the use of VR in health promotion efforts.

Study Selection Procedure

All search results were stored in Zotero, an open-source reference manager, and duplicates were removed. Titles and abstracts were screened first, removing articles that clearly did not match eligibility criteria. Second, full texts of the remaining articles were downloaded to define final eligibility for inclusion. For each step, 2 reviewers conducted the screening independently and compared and discussed these discrepancies until a full consensus was reached.

Data Extraction Process and Synthesis of Results

Data extraction was done by 1 reviewer, who extracted the following items from the included articles: (1) title and authors, (2) goal(s) of the study, (3) design of the study, (4) study sample characteristics, (5) VR device used, (6) main results, and (7) limitations reported by the authors. A second reviewer verified that all data were correctly extracted. Following the data extraction, we conducted a narrative analysis and synthesis of the results. Results and implications of the data extracted from the included studies were discussed by 2 reviewers in relation to the 3 research questions of the scoping review.

Results

Study Selection and Characteristics of Included Studies

The initial search identified 4295 unique articles, which were reduced to 51 eligible articles (see the PRISMA flowchart in Figure 1). Included studies were conducted in Asia (11 studies, 22%), Europe (18 studies, 35%), the Middle East (1 study, 2%), and North America (21 studies, 41%). The total sample size across all studies was 4647 participants, with an average of 91.1 participants per study. Study samples included slightly more women, with 2651 women (53%) and 1958 men (42.7%). The mean age of participants across the studies was 31.6 (SD 5.45) years. Studies primarily included adults, with 29 studies (57%) focused on adults, followed by 14 studies (28%) focused on adolescents, 7 studies (14%) focused on senior adults, and 1 study (2%) focused on children. Specific populations studied included students (7 studies, 29%), people with cognitive impairment (3 studies, 12%), and people with obesity (3 studies, 12%). Other populations studied included former smokers (1 study, 4%), NHS staff (1 study, 4%), parents (2 studies, 8%), smokers (2 studies, 8%), adults who had been in lockdown (1 study, 4%), and unvaccinated adults (1 study, 4%, see Figure 2 for details).

Figure 1. Flowchart of the study selection process following PRISMA guidelines. A total of 4295 articles were initially identified across 3 databases. After removing duplicates and applying eligibility criteria, 51 studies were included. iVR: immersive virtual reality; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses.

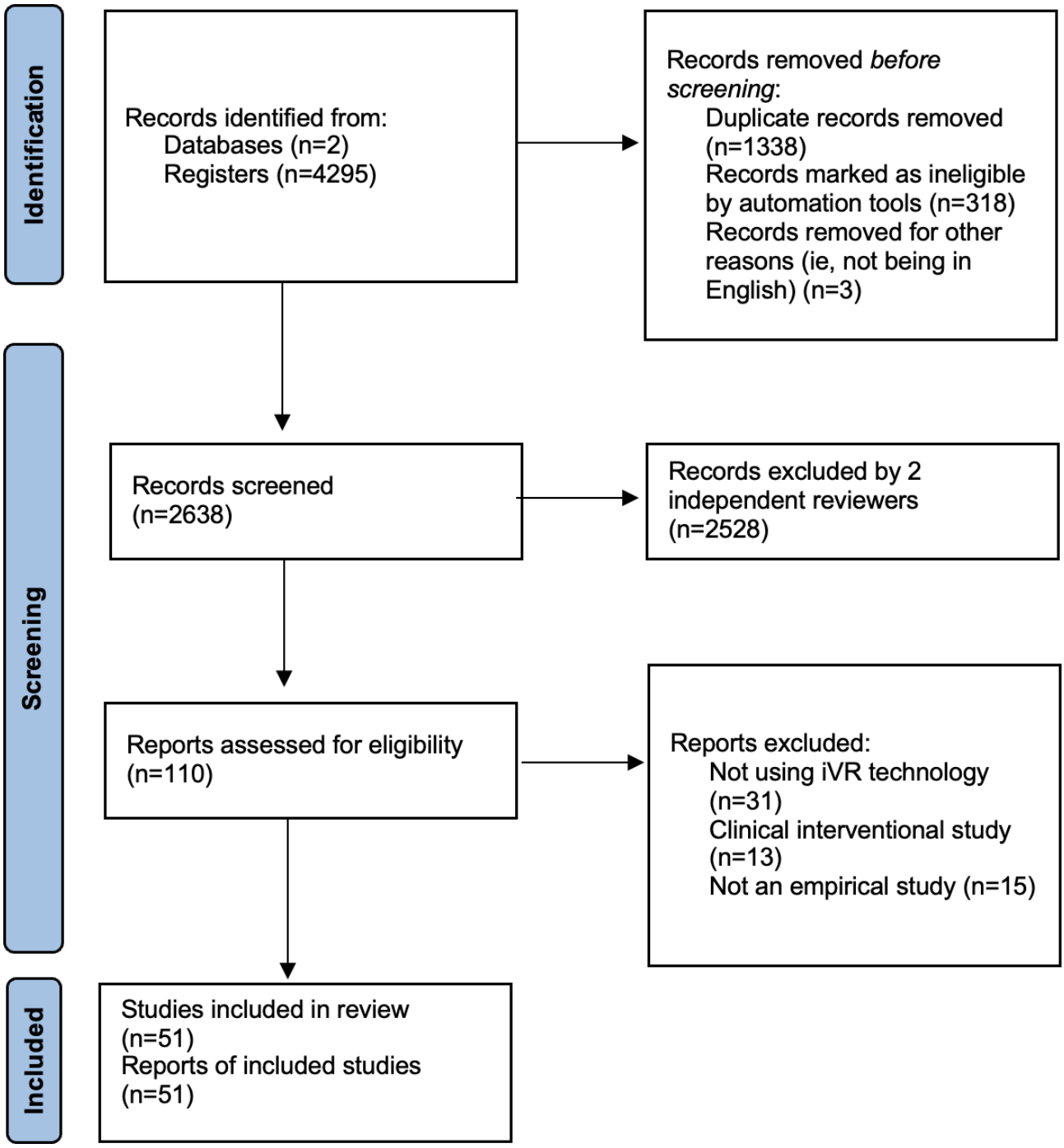
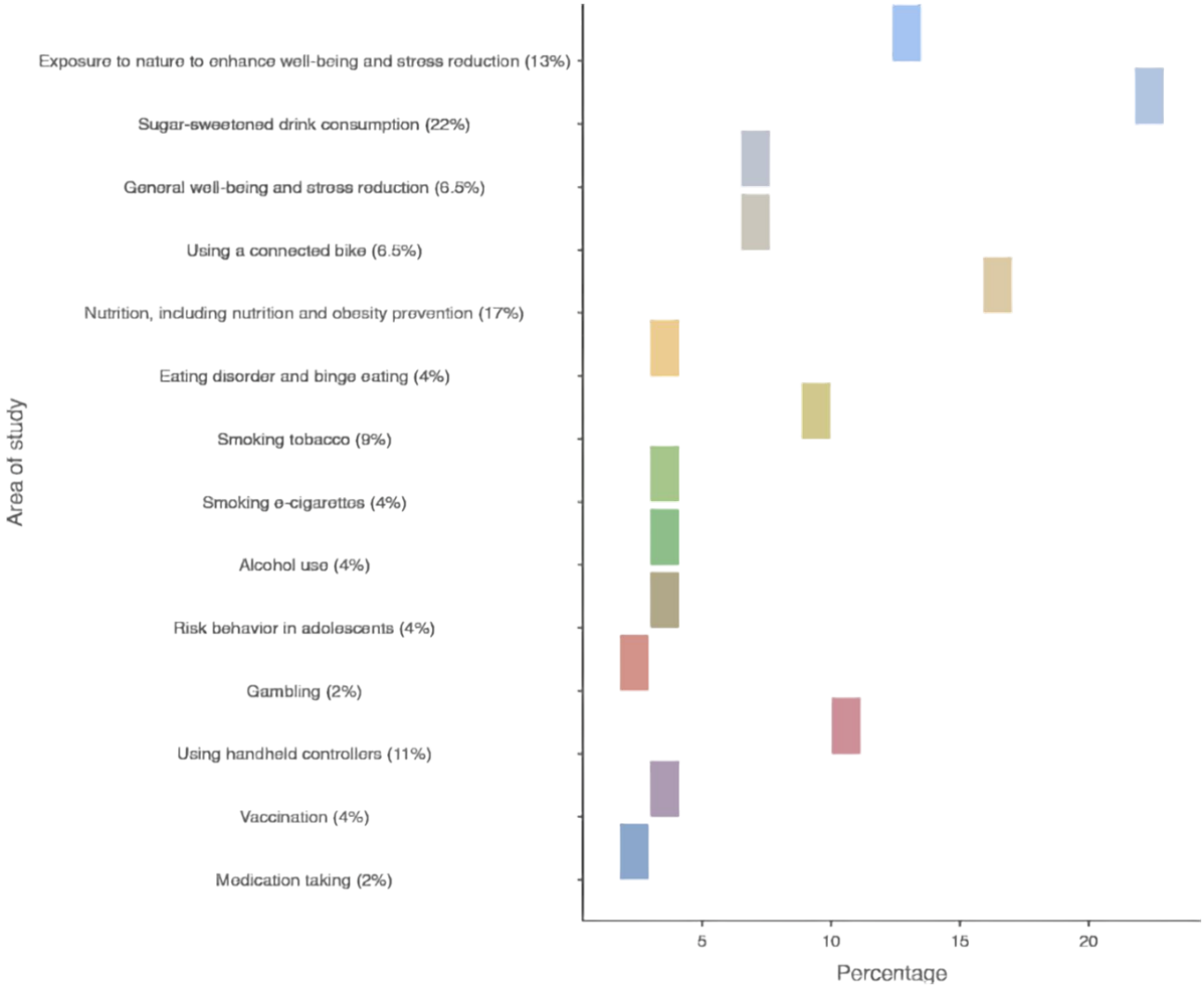


Figure 2. Overview of key outcomes from the intervention studies.



Characteristics of Sources of Evidence

The 51 included studies focused on various health-related topics (Table 1), the most predominant ones being nutrition (17%) and risky behaviors (4%). All studies used HMD, except for Lemieux et al [25], where the device used was not mentioned. HMDs were mainly Oculus (Quest, Go, or Rift, 24%), HTC Vive (17%), or Samsung Gear VR (15%). Almost half of the studies (43%) were coupled with 1 or 2 handheld controllers. Most studies (56%) included an active interaction with the VE by using 1 or 2 handheld controllers or the bodysuit to interact with the VE. About 49% of VR exposure lasted a maximum of 10 minutes, including 22% of studies with under 5 minutes of VR exposure. We estimated an average time of 12.8 (SD 11.1) minutes spent using VR, according to the information given in the articles.

Table . Characteristics of sources of evidence.

Area of study	Studies
Nutrition, including nutrition and obesity prevention (17%)	Blom et al [26]; Isgin-Atici et al [27]; Ledoux et al [28]; Marcum et al [29]; McBride et al [30]; Persky et al [31,32]; Verhulst et al [33]
Eating disorder and binge eating (5%)	Ferrer-Garcia et al [34]; Lemieux et al [25]
Sugar-sweetened drink consumption (10%)	Blom et al [26]; Ledoux et al [28]; Marcum et al [29]; McBride et al [30]
Smoking tobacco (8%)	Borrelli et al [35]; Ferrer-García et al [36]; García-Rodríguez et al [37]; Bonneterre et al [17]
Smoking e-cigarettes (5%)	Weser et al [38,39]
Alcohol use (5%)	Guldager et al [40]; Ma [41]
Risk behavior in adolescents (4%)	Hadley et al [42,43]
Gambling (2%)	Detez et al [44]
Exposure to nature to enhance well-being/stress reduction (14%)	Alyan et al [45]; Beverly et al [46]; Brimelow et al [47,48]; Browning et al [49]; Calogiuri et al [50]
General well-being/stress reduction (10%)	Afifi et al [51]; Adhyaru et al [52]; Kim et al [53]; Riva et al [54]; Ko et al [55]; Kiper et al [56]
Using handheld controllers (7%)	Eisapour et al [57]; Fang and Huang [58]; Farič et al [59]
Using a connected bike (7%)	Bird et al [60]; Zeng et al [61,62]
Vaccination (4%)	Mottelson et al [63]; Nowak et al [64]
Medication-taking (2%)	Niki et al [65]

Results of Individual Sources of Evidence: Detailed Results

Main Identified Research Goals

We identified three main goals for using VR: (1) as a tool to deliver an intervention, with 35 articles focusing on either (1a) pilot testing or testing the feasibility of using VR materials or procedures or (1b) using VR to deliver an actual intervention (eg, skills learning, comparing VR vs other intervention modalities) to test its relative efficacy; (2) as a tool to address fundamental research questions, with 6 studies aimed at recreating ecological settings to address physiological and psychological changes when exposed to certain situations (eg, cravings elicitation); or (3) as an assessment tool, with 5 studies investigating food choices with a food buffet created in VR.

Pilot Studies: Ensuring Usability and Enjoyability

Many studies included in the scoping review were pilot or feasibility studies (about 37%, Table 2) from which we distinguished two main purposes: (1) testing VR usability for future research and seeing how target outcomes are impacted and (2) assessing users' experience with VR. First, researchers found that the use of VR in their methods was rather relevant and reached multiple target outcomes such as reducing stress using a short exposure to nature in VR [45-48,51,52,55], even though exposure durations were relatively short (3-10 minutes). The use of VR was also useful to enhance participants' physical and cognitive activity [66,67]. Finally, some studies were focused on prevention and the major advantage of VR use is its ability to involve participants directly in the preventive message, for example through gaming [68] or skill practice (eg, refusing peer pressure to vape [38]). This resulted in improved knowledge on health topics (eg, on smoking in [69]) and intentions to check for diseases (eg, [20]). It also helped to deliver information in a more traditional preventive way (eg, exposure to a preventive video in an HMD in [35] or a FestLab in [40]). Overall, pilot and feasibility studies, even if conducted on small samples, found VR to be enjoyed and accepted by participants, as well as useful and feasible, and found that it impacted target outcomes (eg, enhanced well-being, increased knowledge). These results occurred whether participants only had a one-time exposure (eg, [53,69]) or sessions over a few weeks (eg, [48]) and were found to be sustained at follow-up when measured (eg, participants reduced their tobacco intake over the month following their participation [35]).

Table . Summary of articles and their classification within the scoping review.

Category and study	Descriptives	iVR[a] details	Objective(s)	Study design	Main conclusions
1a: Pilot or feasibility studies					
Adhyaru and Kemp [52]	n=39; mean age 36.6 (SD 10.3) years; 82% women; health care workers	HMD[b] (Oculus Go); 10 minutes	Explore if exposure to nature in iVR can help health care workers de-stress at work.	Before-after exposure; within-subject	iVR reduced anxiety, anger, and heart rate, and enhanced happiness and relaxation.
Afifi et al [51]	n=50 older adults with cognitive impairments and their family members	Immersive VR[c] system	Assess whether iVR improves quality of life and social interaction for older adults and their family members.	Feasibility study with pre-post assessments	VR improved social interaction and quality of life for both older adults and their families.
Alyan et al [45]	n=20; mean age 21.8 (SD 2.2) years; 50% women; students	HMD (HTC Vive); 5 minutes	Use iVR to reduce stress via a virtual walk in nature.	2 (environment: realistic vs dreamlike); between-subject	iVR reduced stress and enhanced mental well-being.
Beverly et al [46]	n=102; 72% women; health care workers	HMD (Oculus Go/Pico G2); 3 minutes	Explore if cinematic iVR can reduce stress in health care workers.	Before-after exposure; within-subject	iVR reduced stress, independently of previous iVR use or job type.
Bonneterre et al [17]	n=121; mean age 19.6 years; 82.5% female; university students	Sensiks Immersive VR system	Evaluate the impact of VR on memorization, attitudes, and craving responses to anti-tobacco posters.	Randomized controlled trial	VR enhanced memorization of prevention messages.
Borelli et al [35]	n=23; mean age 49.8 (SD 13.3) years; 22% women; adult smokers	HMD (Knoxlabs V2 cardboard); 5 minutes	Examine the feasibility and impact of a smoking cessation intervention during dental cleaning.	2 (video type: smoker ready/not ready to quit) × 3 (time: pre/post/follow-up); within-subject	Feasible and accepted by both smokers and dental care providers.
1b: Interventions					
Ahn [5]	n=73; mean age 20.8 (SD 1.1) years; 82% women; students	HMD (NM); 2 minutes	Test efficacy of preventive messages on sugar and sweetened beverage consumption via avatar embodiment.	2 (pamphlet only vs pamphlet plus iVR) × 2 (tailoring: others vs self) × 3 (time: pre/post/follow-up); between-subject	iVR heightened intentions to limit sugar and sweetened beverage consumption; effects were present at follow-up.
Blom et al [26]	n=99; mean age 30.7 (SD 10.9) years; 60% women; general population	HMD (HTC Vive); ≥3 minutes	Study purchase behaviors in an iVR supermarket.	2 (nudge vs control) × 2 (time pressure: 3 minutes vs no pressure); between-subject	iVR revealed changes in healthy food purchases based on nudge type.
2: Fundamental research					
Chittaro et al [70]	n=105; mean age 21.49 (SD 2.43) years; 90.5% women; students	HMD (Sony HMZ-T1); 5 minutes	Investigate links between iVR and persuasion theory, including inducing mortality salience.	2 (environment: iVR park vs cemetery); between-subject	iVR elicited mortality salience, impacted attitudes, and induced greater physiological reactions than traditional mortality salience manipulations.
Ferrer-Garcia et al [36]	n=25; mean age 29.7 (SD 13.4) years; 32% women; smokers	HMD (5DT HMD 800); time not mentioned	Assess iVR's ability to produce cravings toward tobacco smoking.	Before-during exposure to smoking cues	iVR created cravings, correlated with presence.
3: Assessment tool					
Isgin-Atici et al [27]	n=73; mean age 22.2 (SD 4.1) years; 56% women; students	HMD (HTC Vive); 5 - 25 minutes	Evaluate ease of use and efficiency of a virtual cafeteria.	2 (groups: iVR novices vs experienced); between-subject	iVR was user-friendly and effective regardless of prior VR experience.

Category and study	Descriptives	iVR[a] details	Objective(s)	Study design	Main conclusions
Marcum et al [29]	n=221; mean age 38 (SD 5.6) years; 100% women; mothers with obesity	HMD; time not mentioned	Examine microbehaviors influencing food selection in an iVR buffet.	3 (conditions: food safety control vs behavioral risk information vs family-based risk information); between-subject	iVR enabled dynamic assessment of food choice behaviors.

[a]iVR: immersive virtual reality.

[b]HMD: head-mounted display.

[c]VR: virtual reality.

Second, most participants found VR enjoyable and fun [59,68] and quite easy to use [52]; some were asked to complete a short tutorial [27]. Even older adults were able to manipulate handheld controllers [57], but 1 study reported that the HMD is sometimes heavy for their neck to lift (1 participant dropped out because of this reason [52]). It is worth noting that some of these studies [20,59] involved the targeted population in co-designing the intervention in previous pilot studies, hence not only explicitly ensuring usability [57] but also enhancing users' satisfaction with the intervention. Co-designing an intervention with the targeted population and conducting a first pilot study on a small sample (eg, 12/33) can improve the level of satisfaction and usability of the intervention prototype, albeit ultimate user satisfaction can only be assessed following full-scale deployment of the intervention.

Relative Efficacy of VR Interventions

Interventions (39% [20/51] of the studies included in the review) using VR focused on several targets such as enhancing well-being by simulating a walk in nature (while remaining seated [49,54] or walking on a treadmill [50]) or skill learning and practice on various health topics [40,42]. Some studies were interested in delivering preventive content [40,41,71], other studies used VR's ability to create standardized conditions to test theoretical frameworks (eg, nudge and time pressure on healthy food choice [26,40]), while still others used VR to embody a specific character in order to impact health outcomes [5,33,56].

The key element of most studies included in this group is that they often compared the use of VR with other modalities to deliver an intervention; for example, delivering preventive information in VR versus a 2D screen (eg, [41,54]) or without the use of specific technology (eg, live role-playing with an instructor [64], reading a pamphlet, [61]). Some studies also compared different depths of immersion [50,60,62].

When comparing the relative efficacy of VR with other modalities, mixed results were found. For example, even though participants exercising using VR experienced an attentional shift from exercising, meaning that individuals were usually distracted and entertained by the VR setting, leading them to actually enjoy physical exercise, it was not always sufficient to obtain greater physical involvement when compared to nonimmersive physical activities [25,58,60,62]. However, some studies found no difference in outcomes between the use of VR and 2D screens [54,71], and other studies even found that a virtual walk remained less efficient than a real walk in nature for mood enhancement [49]. Some studies, using VR only, also found no impact of VR prevention interventions on target outcomes (eg, no change in physical self-perception when using VR to prevent eating disorders [39], no increased knowledge on alcohol [40]). Still, we note that VR was a great tool to induce changes in knowledge and intentions to adopt a behavior (eg, vaccination intention [61], smoking e-cigarettes [44]) and for skill practice [42].

A few recent studies [44,54] investigated the use of VR outside of the laboratory, recruiting participants who own VR devices at home. Portable VR devices have become more affordable, resulting in individuals being able to use them potentially anywhere and be autonomously engaged with VR-based interventions. Furthermore, both studies resulted in an improvement in the target outcomes (reduction of psychological distress [54], increase in vaccination [44]).

Overall, VR is impactful; it can create precise and standardized experimental situations (eg, embodying an obese or weight-gaining avatar [5,33]), and it is especially practical for skill practice and sometimes for physical activity. VR-based interventions have shown a higher degree of attendance in intervention sessions (ie, adherence) than the same intervention done without the use of VR [64]. However, when VR is only used to deliver information without leveraging its specific characteristics, such as immersivity and active use of the device (ie, interacting with the VE via a game [61]), it has often been found to have similar efficacy as more traditional ways to deliver information (eg, 2D screens).

VR to Address Fundamental Health Research Questions: A Tool to Recreate Ecological Settings in the Lab

VR can recreate real-life situations in laboratories and has been used across different domains, such as gambling [36], tobacco cravings [28,37], and food cravings [34,70], as well as for mimicking specific situations inducing certain psychological states, such as mortality salience (eg, [29]). In all studies, exposure to specific cues (eg, food items, cemetery, individuals smoking) or situations (eg, being in a pub, gambling on a slot machine) elicited both physiological (eg, increased heart rate, arousal) and psychological (eg, self-reported craving) changes, whether individuals were actively (ie, interacting with the VE) or passively (ie, watching visual content) using the VR device, suggesting that the highly immersive characteristics of VR are effective at eliciting an emotional response.

However, only 1 study compared eliciting cravings using VR versus other types of devices [34], indicating VR is not better suited to trigger a craving response than 2D pictures. It might

be possible that this null effect was due to the passive use of VR in this specific study, as interacting with a cue in VR has been found to enhance cravings [28].

VR as an Assessment Tool in Health-Related Interventions

A total of 5 studies used VR as an assessment tool in the field of nutrition by recreating a virtual buffet displaying food [27,30-32,65], where participants' task was to collect a plate of food. Participants found the VR food buffet easy to use, independently of whether they already used a VR device in the past [27]. In this context, VR allows researchers to study precisely how many items and types of food were selected and in which quantity, enabling them to calculate the total calories contained in each plate more easily. It also helped to display to participants a standardized food buffet with diverse food items without constraints from a real food buffet (eg, expiration dates, flexibility in food types, reduced costs).

Study Limitations

The Necessity to Adapt the Use of VR to Experimental Needs

The use of VR, whether for applied or fundamental research, has shown some limitations, mainly related to the study methodology and VR technology itself (eg, cybersickness, notably in [50]). First, a majority of included studies suffered from either small sample sizes (eg, 10 participants in [65], 6 in [57]) or nonrepresentative samples (eg, students in [29], healthy and active young individuals in [58]), limiting the validity and generalizability of results. Second, the quality of the experimental designs was sometimes limited (eg, semiexperimental design with pre-post comparisons) because of a lack of a proper control condition or not conducting a rigorous randomized controlled trial [20,46]. Short-term follow-up or the lack of a follow-up altogether was also mentioned as a limiting factor in numerous studies [41,42].

Lack of Systematic Assessment of VR's Main Characteristics: Presence, Immersion, and Cybersickness

VR's effects, especially persuasive effects, seem to come from its ability to enhance presence, which is the feeling of being there during a VR experience. Hence, participants act similarly to real life in the VE because they are fully immersed in their interaction with it. The level of presence experienced by users can impact targeted variables in the intervention; participants who felt more present in the VE showed stronger positive effects on persuasion-related outcomes (eg, attitudes toward vaccination and intention to get vaccinated [61]; higher presence resulted in more reported cravings for tobacco in [37]). However, presence is rarely measured as a moderator or covariate across studies despite its potential impact on outcomes. The same applies to immersion, which was not measured across studies, despite studies often comparing different intervention modalities of varying degrees of immersion (eg, VR versus 2D screen). VR is not the only technology able to generate presence; narrative, videos, or nonimmersive VR can too [14]. Not measuring immersion or presence across different modalities limits the understanding of VR's role in driving effects on the target outcomes.

Finally, cybersickness was rarely measured across studies despite its potential negative effect on user experience and, in turn, target outcomes. Some studies, notably the ones focusing on physical activity, measured cybersickness and found that it can completely erase the positive effects of using VR (eg, walking on a treadmill while wearing a VR device led to cybersickness, which diminished the positive effects of being exposed to nature compared to the other condition, [50]). Participants who felt symptoms of cybersickness believed that it impacted their experience [59], sometimes to the point they had to drop out of the experiment [62].

Discussion

Principal Findings

This scoping review identified 51 studies published over the past 14 years that explored the use of VR in health promotion and prevention psychology. Our findings revealed three primary applications of VR: (1) as a tool to deliver interventions, either in feasibility testing or actual implementation; (2) as a means to address fundamental research questions; and (3) as an assessment tool for health-related outcomes. Although VR shows significant promise in creating immersive and engaging interventions, our review highlights the variability in effectiveness and common challenges such as small sample sizes, short follow-up periods, and limited methodological standardization.

VR technology use for health promotion and prevention research is relatively recent, with studies in this review indicating its potential as a promising tool to deliver and assess interventions. For instance, VR was effective in simulating realistic scenarios to engage participants in skills-based learning and decision-making tasks, such as risk-reduction behaviors [42,61]. VR allows researchers to create safe, ecological, and standardized VEs in which it is possible to deliver and evaluate health promotion and preventive interventions [42]; recreate situations or environments that can elicit strong emotional, physiological, behavioral, or psychological responses (eg, mortality salience [29]); and assess outcomes (eg, cravings, food choices) with a multimeasure approach included in VR technologies (eg, psychological, physiological, and behavioral measures). This scoping review identified 51 studies concerning the use of VR technology in the field of health promotion and prevention psychology published within the past 14 years. We mapped (1) the goals and research questions addressed through the use of VR in this field, (2) its effects in the identified areas, and (3) its main challenges or limitations. We identified three main applications of VR in this field: (1) as a tool to deliver an intervention, either (1a) pilot or feasibility testing VR materials or procedures or (1b) using VR to deliver an actual intervention (eg, skills learning, comparing VR vs other intervention modalities) to test its relative efficacy; (2) as a tool to address fundamental research questions; and (3) as an assessment tool.

Comparison to Prior Work

Due to the relative novelty of VR in this field, only 51 eligible studies were published in the past 14 years. Research so far has mostly focused on feasibility or pilot studies, aimed at testing the ability of VR to be integrated into interventions [69], with

a minority of studies focusing on answering fundamental research questions through the use of VR [28]. Most studies employed semiexperimental designs without a control or comparison group and often had a short or no follow-up, limiting the validity and generalizability of results. Studies also included relatively small samples and were often nonrepresentative of the general population (eg, students). However, as the use of VR in the field of health promotion and prevention is still in its infancy, it appears natural to see a stronger focus on pilot or feasibility studies in the published literature.

Strength and Limitations of the Scoping Review

When considering whether VR is effective in health prevention, it should first be noted that the effectiveness of VR interventions was variable. For instance, while some studies indicated that VR could enhance user engagement and motivation [59], others found no significant differences in outcomes compared to traditional methods [71]. This highlights the need for further research to clarify the conditions under which VR is most effective. This scoping review showed that sometimes VR use is not systematically more effective in achieving target outcomes than its nonimmersive equivalents [41,54,71]; we supposed that to be more effective, VR should be used for its specific immersive characteristics, such as gamification or embodiment, which directly involve the user. For example, skills practice in VR was more effective than role-playing in real life to learn about risk behaviors and ways to avoid them (eg, buying condoms for safer sex) due to VR scenarios' ability to recreate a situation that is realistic, induce emotional changes in the user as the scenario goes on, and finally, make the user have a real first-person experience [42]. Similar results appeared in [61], in which VR was used to represent a vaccination intervention to stop flu spread (ie, participants used handheld controllers to actively send immune cells to prevent flu transmission), whereas in other conditions, participants were just passively watching (a video on a 2D screen or a pamphlet).

It is important to consider the limitations of this review when interpreting the findings. First, as the use of VR in health promotion and prevention psychology is a relatively recent phenomenon, our literature search focused on the last 14 years (2010 - 2024). This resulted in the inclusion of 51 eligible articles, which may have excluded earlier or less accessible studies. However, the majority of included studies (63%) were published between 2020 and 2024, reflecting the increasing affordability and accessibility of VR technology for research in recent years. Therefore, the likelihood of missing pivotal studies is low. Second, the search strategy did not include gray literature, which may have reduced the total number of eligible articles and introduced publication bias by excluding studies with nonsignificant or null results (the file drawer effect). To address this gap, future reviews should consider including gray literature to provide a more comprehensive overview of the field. Third, some studies lacked sufficient reporting of critical aspects such as sample characteristics (eg, size and demographics) and details of VR implementation (eg, exposure duration, type of VR technology used). This limited our ability to draw broad conclusions about the efficacy and applicability of VR in this area. Addressing these reporting gaps in future research will improve the comparability and quality of evidence in this rapidly evolving area of study. Fourth, although our literature search was updated during the initial revision, which was completed just a few weeks prior to this submission, we recognize that VR research is advancing rapidly. It is therefore possible that new studies may emerge shortly after the conclusion of our search period, which may influence the results of future reviews. To address this, future updates could consider conducting more frequent searches or establishing a continuous review process to ensure that all emerging data are included in real time. However, we are confident that this review accurately reflects the state of the literature as of our latest search.

Perspectives and Future Research Directions

Standardization of Designs and Replication

Although the results of our scoping review suggest that VR has potential as a tool for health promotion, the field is still in its infancy. Many studies in this area are limited by small sample sizes, short follow-up periods, and inadequate experimental control. Replication is essential to strengthen the reliability and validity of these findings [72,73]. Replication of these studies in diverse populations and settings will help confirm the generalizability of the findings and identify any boundary conditions, such as differences in user demographics, technology exposure, or the specific health behaviors targeted [74,75]. In addition, replication can shed light on the mechanisms underlying the effectiveness of VR interventions, which may vary depending on the context and population studied. Therefore, further replication is essential not only to solidify current evidence, but also to ensure that VR interventions are applicable and effective across a wide range of health promotion and primary prevention efforts. Replicating existing results to increase the amount and quality of empirical evidence supporting the use and benefit of VR in this field is needed. For example, in this scoping review, we saw that individuals showed an increased knowledge regarding health-related topics [20,69] or changed their behavioral intentions [44,61] when exposed to a VR intervention. However, not all studies provided evidence to fully support these claims [47,63], in addition to the lack of any perspective on how long these effects last or if they are applicable to less specific populations. Therefore, a focus on study replication can strengthen the advancement of research in this field and at the same time prevent a replication crisis, as observed in other fields of behavioral sciences and medicine [76]. There is also a critical need for future research to employ longitudinal study designs. Long-term follow-up is particularly important in preventive psychology, where sustained behavior change and long-term health outcomes are key indicators of success.

VR vs Nonimmersive Apparatus

The effectiveness of VR compared to nonimmersive interventions, such as 2D presentations, remains controversial. Evidence from the studies included in this review showed mixed results. Although some studies reported that VR interventions increased engagement and enjoyment, others found no significant differences in outcomes compared to nonimmersive methods [59,71]. The immersive features of VR, such as gamification and embodiment, appear to be particularly effective in scenarios that require active user involvement. For example,

participants who practiced risk-avoidance skills in VR showed better retention than those who used real-life role-playing [42]. Similarly, the use of VR in interactive scenarios, such as vaccination education, showed higher levels of engagement than passive modalities such as 2D videos or pamphlets [61]. However, studies have also shown that VR does not always outperform traditional methods in terms of physical activity or knowledge acquisition. This variability highlights the need for future research to clarify the specific contexts in which the immersive qualities of VR are most effective. Systematic assessment of key mechanisms such as presence and immersion could help determine whether VR's effectiveness is primarily due to its immersive nature or to other factors such as interactivity or novelty.

Assessing presence and immersion is crucial for understanding the mechanisms underlying VR and its effects [14]. Evaluating the feeling of presence helps determine the extent to which participants are psychologically immersed in VEs and allows for the identification and correction of potential errors in the VE that could influence presence and, consequently, the effectiveness of VR-based interventions or content. Additionally, addressing such errors can prevent cybersickness and ensure the smooth execution of experiments [50]. Measuring presence and immersion provides valuable insights into individuals' capacity to engage with VR compared to nonimmersive interventions and helps identify how these factors correlate with target outcomes.

Set Up for Success

Conducting feasibility or pilot studies to test the VR procedure and VEs is recommended. As shown in the scoping review, evaluating the enjoyability, usability, and safety of the procedure can be very helpful. Finally, co-designing the VR-based intervention with participants from the targeted population can enhance the relevance, validity, and user experience with the intervention itself. Cocreating a procedure with participants could induce a bias in their judgment, making them judge the intervention more positively than it actually is. Pilot testing with different groups of participants is recommended to validate the final design.

Make It Simple and Clear for Participants

VR studies are attractive to participants (eg, higher attendance for intervention sessions than the non-VR condition in [19,42,77]), but they can be complex to follow all the way through (ie, risk of cognitive overload, fatigue [78]). When designing studies using VR, keeping them as simple and short as possible will minimize participant burden and fatigue. It is also highly possible that most participants have never experienced VR before, so making sure they understand how to move and interact with the environment at first is necessary. If possible, we recommend doing a short tutorial on how to use the controllers or putting the participant in a tutorial VE before the experimental procedure. The participants can then fully concentrate on what is happening in the VR rather than think about how to interact with the VE.

Conclusion

This scoping review provides an overview of VR's emerging role in health promotion and prevention psychology, highlighting its potential to create immersive and engaging interventions. Although VR has shown promise in delivering health interventions and answering fundamental research questions, its effectiveness remains variable, and many studies are limited by methodological constraints. Future research should prioritize replication, longitudinal designs, and standardized methodologies to strengthen the evidence base and realize the full potential of VR in this field.

Acknowledgments

We would like to acknowledge the support of our respective institutions: University Paris Nanterre, Université Clermont Auvergne, Université Rouen Normandie, and Erasmus University Rotterdam. We also thank our colleagues for their insightful discussions and suggestions while conducting this research. This study has been funded by an IresP-INCa grant (number 19II028-00).

Data Availability

This scoping review is based on a synthesis of publicly available research articles. The datasets analyzed during this study are derived from published sources that are referenced within the manuscript. No new datasets were generated.

Authors' Contributions

SB contributed to the conceptualization, formal analysis, methodology, and writing of the original draft of the manuscript. OZ was involved in the conceptualization, formal analysis, methodology, writing of the original draft, and reviewing and editing the manuscript, while also providing supervision throughout the project. MB participated in the conceptualization, methodology, and writing of the original draft, and offered supervision during the research process.

Multimedia Appendix 1
Summary of Studies Using Virtual Reality for Health-Related Behavioral Interventions.
[DOCX File, 37 KB - xr_v2i1e49923_app1.docx]

Checklist 1

PRISMA-ScR checklist.
[PDF File, 326 KB - xr_v2i1e49923_app2.pdf]

References

1. Green LW, Kreuter M. Health Program Planning: An Educational and Ecological Approach, 4th edition: McGraw Hill Higher Education; 2004.
2. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
3. Kennedy W, Fruin R, Lue A, Logan SW. Using ecological models of health behavior to promote health care access and physical activity engagement for persons with disabilities. J Patient Exp 2021;8:23743735211034031. [doi: 10.1177/23743735211034031] [Medline: 34350340]
4. Schultheis MT, Rizzo AA. The application of virtual reality technology in rehabilitation. Rehab Psychol 2001;46(3):296-311. [doi: 10.1037//0090-5550.46.3.296]
5. Ahn SJG. Incorporating immersive virtual environments in health promotion campaigns: a construal level theory approach. Health Commun 2015;30(6):545-556. [doi: 10.1080/10410236.2013.869650] [Medline: 24991725]
6. Bailey JO, Bailenson JN, Flora J, Armel KC, Voelker D, Reeves B. The impact of vivid messages on reducing energy consumption related to hot water use. Env Behav 2015 Jun;47(5):570-592. [doi: 10.1177/0013916514551604]
7. Bailenson J. Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do: WW Norton & Company; 2018. URL: https://psycnet.apa.org/record/2019-04570-000 [accessed 2024-12-15]
8. Rizzo A, Bouchard S. Virtual Reality for Psychological and Neurocognitive Interventions: Springer; 2019.
9. McCall C, Blascovich J. How, when, and why to use digital experimental virtual environments to study social behavior. Soc & Pers Psych 2009 Sep;3(5):744-758. [doi: 10.1111/j.1751-9004.2009.00195.x]
10. Yaremych HE, Persky S. Tracing physical behavior in virtual reality: a narrative review of applications to social psychology. J Exp Soc Psychol 2019 Nov;85:103845. [doi: 10.1016/j.jesp.2019.103845] [Medline: 32831397]
11. Cummings JJ, Bailenson JN. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Med Psychol 2016 Apr 2;19(2):272-309. [doi: 10.1080/15213269.2015.1015740]
12. Parsons TD, Gaggioli A, Riva G. Virtual reality for research in social neuroscience. Brain Sci 2017 Apr 16;7(4):42. [doi: 10.3390/brainsci7040042] [Medline: 28420150]
13. Ahn SJG, Hahm JM, Johnsen K. Feeling the weight of calories: using haptic feedback as virtual exemplars to promote risk perception among young females on unhealthy snack choices. Med Psychol 2019;22(4):626-652. [doi: 10.1080/15213269.2018.1492939] [Medline: 32863775]
14. Pianzola F, Riva G, Mantovani F, Kukkonen K. Presence, flow, and narrative absorption: an interdisciplinary theoretical exploration with a new spatiotemporal integrated model based on predictive processing. PsyArXiv. Preprint posted online on Jul 2, 2020. [doi: 10.31234/osf.io/uzrcm]
15. Yao S, Kim G. The effects of immersion in a virtual reality game: presence and physical activity. Presented at: 21st International Conference on Human-Computer Interaction; Jul 26-31, 2019; Orlando, FL, USA p. 234-242. [doi: 10.1007/978-3-030-22602-2_18]
16. Schwind V, Knierim P, Haas N, Henze N. Using presence questionnaires in virtual reality. Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 2, 2019; Glasgow, Scotland, UK p. 1-12. [doi: 10.1145/3290605.3300590]
17. Bonneterre S, Zerhouni O, Boffo M. The influence of billboard-based tobacco prevention posters on memorization, attitudes, and craving: immersive virtual reality study. J Med Internet Res 2024 Jul 9;26:e49344. [doi: 10.2196/49344] [Medline: 38980707]
18. Schmälzle R, Lim S, Cho HJ, Wu J, Bente G. Examining the exposure-reception-retention link in realistic communication environments via VR and eye-tracking: the VR billboard paradigm. PLoS One 2023;18(11):e0291924. [doi: 10.1371/journal.pone.0291924] [Medline: 38033032]
19. Riva G, Baños RM, Botella C, Mantovani F, Gaggioli A. Transforming experience: the potential of augmented reality and virtual reality for enhancing personal and clinical change. Front Psychiatry 2016;7:164. [doi: 10.3389/fpsyt.2016.00164] [Medline: 27746747]
20. Saab MM, Landers M, Cooke E, Murphy D, Davoren M, Hegarty J. Enhancing men's awareness of testicular disorders using a virtual reality intervention. Nurs Res 2018;67(5):349-358. [doi: 10.1097/NNR.0000000000000303] [Medline: 30059354]
21. Hamilton-Giachritsis C, Banakou D, Garcia Quiroga M, Giachritsis C, Slater M. Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. Sci Rep 2018 Feb 14;8(1):2975. [doi: 10.1038/s41598-018-21036-2] [Medline: 29445183]
22. Tricco AC, Lillie E, Zarin W, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 9;16:15. [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
23. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]

24. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890. [doi: 10.1136/bmj.l6890] [Medline: 31948937]
25. Lemieux V, Monthuy-Blanc J, Corno G. Application of virtual reality in an ED primary prevention context: an exploratory study. Annu Rev CyberTher Telemed 2020;18:207-211 [FREE Full text]
26. Blom S, Gillebaart M, De Boer F, van der Laan N, De Ridder DTD. Under pressure: nudging increases healthy food choice in a virtual reality supermarket, irrespective of system 1 reasoning. Appetite 2021 May 1;160:105116. [doi: 10.1016/j.appet.2021.105116] [Medline: 33450297]
27. Isgin-Atici K, Ozkan A, Celikcan U, et al. Usability study of a novel tool: the virtual cafeteria in nutrition education. J Nutr Educ Behav 2020 Nov;52(11):1058-1065. [doi: 10.1016/j.jneb.2020.08.001] [Medline: 33168205]
28. Ledoux T, Nguyen AS, Bakos-Block C, Bordnick P. Using virtual reality to study food cravings. Appetite 2013 Dec;71:396-402. [doi: 10.1016/j.appet.2013.09.006] [Medline: 24055758]
29. Marcum CS, Goldring MR, McBride CM, Persky S. Modeling dynamic food choice processes to understand dietary intervention effects. Ann Behav Med 2018 Feb 17;52(3):252-261. [doi: 10.1093/abm/kax041] [Medline: 29538666]
30. McBride CM, Persky S, Wagner LK, Faith MS, Ward DS. Effects of providing personalized feedback of child's obesity risk on mothers' food choices using a virtual reality buffet. Int J Obes (Lond) 2013 Oct;37(10):1322-1327. [doi: 10.1038/ijo.2013.87] [Medline: 23736369]
31. Persky S, Ferrer RA, Klein WMP, et al. Effects of fruit and vegetable feeding messages on mothers and fathers: interactions between emotional state and health message framing. Ann Behav Med 2019 Aug 16;53(9):789-800. [doi: 10.1093/abm/kay088] [Medline: 30395145]
32. Persky S, Yaremych HE, Goldring MR, Ferrer RA, Rose MK, Hollister BM. Investigating the efficacy of genetic, environmental, and multifactorial risk information when communicating obesity risk to parents of young children. Ann Behav Med 2021 Jul 22;55(8):720-733. [doi: 10.1093/abm/kaaa103] [Medline: 33196082]
33. Verhulst A, Normand JM, Lombart C, Sugimoto M, Moreau G. Influence of being embodied in an obese virtual body on shopping behavior and products perception in VR. Front Robot AI 2018;5:113. [doi: 10.3389/frobt.2018.00113] [Medline: 33500992]
34. Ferrer-Garcia M, Gutiérrez-Maldonado J, Agliaro-López M, Lobera-Espi X, Pla J, Vilalta-Abella F. Validation of VR-based software for binge eating treatment: preliminary data. Annu Rev CyberTher Telemed 2014;12:146-150. [doi: 10.3233/978-1-61499-401-5-146]
35. Borrelli B, Rueras N, Jurasic M. Delivery of a smoking cessation induction intervention via virtual reality headset during a dental cleaning. Transl Behav Med 2021 Feb 11;11(1):182-188. [doi: 10.1093/tbm/ibz144] [Medline: 31665495]
36. Ferrer-García M, García-Rodríguez O, Gutiérrez-Maldonado J, Pericot-Valverde I, Secades-Villa R. Efficacy of virtual reality in triggering the craving to smoke: its relation to level of presence and nicotine dependence. Annu Rev CyberTher Telemed 2010;8:99-102. [doi: 10.3233/978-1-60750-561-7-123]
37. García-Rodríguez O, Weidberg S, Gutiérrez-Maldonado J, Secades-Villa R. Smoking a virtual cigarette increases craving among smokers. Addict Behav 2013 Oct;38(10):2551-2554. [doi: 10.1016/j.addbeh.2013.05.007] [Medline: 23793042]
38. Weser VU, Duncan LR, Pendergrass TM, Fernandes CS, Fiellin LE, Hieftje KD. A quasi-experimental test of a virtual reality game prototype for adolescent e-cigarette prevention. Addict Behav 2021 Jan;112:106639. [doi: 10.1016/j.addbeh.2020.106639] [Medline: 32920455]
39. Weser VU, Duncan LR, Sands BE, et al. Evaluation of a virtual reality e-cigarette prevention game for adolescents. Addict Behav 2021 Nov;122:107027. [doi: 10.1016/j.addbeh.2021.107027] [Medline: 34225030]
40. Guldager JD, Kjær SL, Grittner U, Stock C. Efficacy of the virtual reality intervention VR FestLab on alcohol refusal self-efficacy: a cluster-randomized controlled trial. Int J Environ Res Public Health 2022 Mar 10;19(6):3293. [doi: 10.3390/ijerph19063293] [Medline: 35328977]
41. Ma Z. The use of immersive stories to influence college students' attitudes and intentions related to drinking and driving. J Am Coll Health 2022 Oct;70(7):2042-2049. [doi: 10.1080/07448481.2020.1842418] [Medline: 33151847]
42. Hadley W, Houck CD, Barker DH, et al. Eliciting affect via immersive virtual reality: a tool for adolescent risk reduction. J Pediatr Psychol 2014 Apr;39(3):358-368. [doi: 10.1093/jpepsy/jst092] [Medline: 24365699]
43. Hadley W, Houck C, Brown LK, Spitalnick JS, Ferrer M, Barker D. Moving beyond role-play: evaluating the use of virtual reality to teach emotion regulation for the prevention of adolescent risk behavior within a randomized pilot trial. J Pediatr Psychol 2019 May 1;44(4):425-435. [doi: 10.1093/jpepsy/jsy092] [Medline: 30551157]
44. Detez L, Greenwood LM, Segrave R, et al. A psychophysiological and behavioural study of slot machine near-misses using immersive virtual reality. J Gambl Stud 2019 Sep;35(3):929-944. [doi: 10.1007/s10899-018-09822-z] [Medline: 30684139]
45. Alyan E, Combe T, Awang Rambli DR, Sulaiman S, Merienne F, Muhaiyuddin NDM. The influence of virtual forest walk on physiological and psychological responses. Int J Environ Res Public Health 2021 Oct 29;18(21):11420. [doi: 10.3390/ijerph182111420] [Medline: 34769937]
46. Beverly E, Hommema L, Coates K, et al. A tranquil virtual reality experience to reduce subjective stress among COVID-19 frontline healthcare workers. PLoS ONE 2022;17(2):e0262703. [doi: 10.1371/journal.pone.0262703] [Medline: 35139092]

47. Brimelow RE, Dawe B, Dissanayaka N. Preliminary research: virtual reality in residential aged care to reduce apathy and improve mood. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):165-170. [doi: 10.1089/cyber.2019.0286] [Medline: 31829729]
48. Brimelow RE, Thangavelu K, Beattie R, Dissanayaka NN. Feasibility of group-based multiple virtual reality sessions to reduce behavioral and psychological symptoms in persons living in residential aged care. J Am Med Dir Assoc 2022 May;23(5):831-837. [doi: 10.1016/j.jamda.2021.07.026] [Medline: 34454923]
49. Browning M, Mimnaugh KJ, van Riper CJ, Laurent HK, LaValle SM. Can simulated nature support mental health? Comparing short, single-doses of 360-degree nature videos in virtual reality with the outdoors. Front Psychol 2019;10:2667. [doi: 10.3389/fpsyg.2019.02667] [Medline: 32010003]
50. Calogiuri G, Litleskare S, Fagerheim KA, Rydgren TL, Brambilla E, Thurston M. Experiencing nature through immersive virtual environments: environmental perceptions, physical engagement, and affective responses during a simulated nature walk. Front Psychol 2017;8:2321. [doi: 10.3389/fpsyg.2017.02321] [Medline: 29410635]
51. Afifi T, Collins N, Rand K, et al. Using virtual reality to improve the quality of life of older adults with cognitive impairments and their family members who live at a distance. Health Commun 2023 Oct;38(9):1904-1915. [doi: 10.1080/10410236.2022.2040170] [Medline: 35253531]
52. Adhyaru JS, Kemp C. Virtual reality as a tool to promote wellbeing in the workplace. D Health 2022;8:20552076221084473. [doi: 10.1177/20552076221084473] [Medline: 35284084]
53. Kim J, Jung YH, Shin YB, et al. Development and validation of a virtual reality-based training program for promoting subjective well-being. Psychiatry Investig 2020 Dec;17(12):1207-1215. [doi: 10.30773/pi.2020.0311] [Medline: 33301665]
54. Riva G, Bernardelli L, Castelnuovo G, et al. A virtual reality-based self-help intervention for dealing with the psychological distress associated with the COVID-19 lockdown: an effectiveness study with a two-week follow-up. Int J Environ Res Public Health 2021 Aug 2;18(15):8188. [doi: 10.3390/ijerph18158188] [Medline: 34360479]
55. Ko H, Lim HJ, Park J, et al. Subjective experience on virtual reality-assisted mental health promotion program. Psychiatry Investig 2024 Apr;21(4):380-386. [doi: 10.30773/pi.2023.0286] [Medline: 38695045]
56. Kiper P, Przysiężna E, Cieślik B, et al. Effects of immersive virtual therapy as a method supporting recovery of depressive symptoms in post-stroke rehabilitation: randomized controlled trial. Clin Interv Aging 2022;17:1673-1685. [doi: 10.2147/CIA.S375754] [Medline: 36447623]
57. Eisapour M, Cao S, Boger J. Participatory design and evaluation of virtual reality games to promote engagement in physical activity for people living with dementia. J Rehabil Assist Technol Eng 2020;7:2055668320913770. [doi: 10.1177/2055668320913770] [Medline: 32499921]
58. Fang YM, Huang YJ. Comparison of the usability and flow experience of an exercise promotion virtual reality programme for different age groups. Behav Inf Technol 2021 Sep 10;40(12):1250-1264. [doi: 10.1080/0144929X.2021.1938680]
59. Farič N, Smith L, Hon A, et al. A virtual reality exergame to engage adolescents in physical activity: mixed methods study describing the formative intervention development process. J Med Internet Res 2021 Feb 4;23(2):e18161. [doi: 10.2196/18161] [Medline: 33538697]
60. Bird JM, Karageorghis CI, Baker SJ, Brookes DA, Nowicky AV. Ready Exerciser One: effects of music and virtual reality on cycle ergometer exercise. Br J Health Psychol 2021 Feb;26(1):15-32. [doi: 10.1111/bjhp.12445] [Medline: 32538512]
61. Zeng N, Liu W, Pope ZC, McDonough DJ, Gao Z. Acute effects of virtual reality exercise biking on college students' physical responses. Res Q Exerc Sport 2022 Sep;93(3):633-639. [doi: 10.1080/02701367.2021.1891188] [Medline: 34663191]
62. Zeng N, Pope Z, Gao Z. Acute effect of virtual reality exercise bike games on college students' physiological and psychological outcomes. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):453-457. [doi: 10.1089/cyber.2017.0042] [Medline: 28715263]
63. Mottelson A, Vandeweerdt C, Atchapero M, et al. A self-administered virtual reality intervention increases COVID-19 vaccination intention. Vaccine 2021 Nov 5;39(46):6746-6753. [doi: 10.1016/j.vaccine.2021.10.004] [Medline: 34654579]
64. Nowak GJ, Evans NJ, Wojdynski BW, et al. Using immersive virtual reality to improve the beliefs and intentions of influenza vaccine avoidant 18-to-49-year-olds: considerations, effects, and lessons learned. Vaccine 2020 Jan 29;38(5):1225-1233. [doi: 10.1016/j.vaccine.2019.11.009] [Medline: 31806533]
65. Niki K, Yasui M, Iguchi M, Isono T, Kageyama H, Ueda M. A pilot study to develop a new method of assisting children in taking their medication by using immersive virtual reality. Biol Pharm Bull 2021;44(2):279-282. [doi: 10.1248/bpb.b20-00646] [Medline: 33518682]
66. Polechoński J, Nierwińska K, Kalita B, Wodarski P. Can physical activity in immersive virtual reality be attractive and have sufficient intensity to meet health recommendations for obese children? A pilot study. Int J Environ Res Public Health 2020 Nov 1;17(21):8051. [doi: 10.3390/ijerph17218051] [Medline: 33139604]
67. Sakhare A, Stradford J, Ravichandran R, et al. Simultaneous exercise and cognitive training in virtual reality phase 2 pilot study: impact on brain health and cognition in older adults. Brain Plast 2021;7(2):111-130. [doi: 10.3233/BPL-210126] [Medline: 34868877]
68. Horsham C, Dutton-Regester K, Antrobus J, et al. A virtual reality game to change sun protection behavior and prevent cancer: user-centered design approach. JMIR Serious Games 2021 Mar 25;9(1):e24652. [doi: 10.2196/24652] [Medline: 33764308]

69. Guo JL, Hsu HP, Lai TM, Lin ML, Chung CM, Huang CM. Acceptability evaluation of the use of virtual reality games in smoking-prevention education for high school students: prospective observational study. J Med Internet Res 2021 Sep 28;23(9):e28037. [doi: 10.2196/28037] [Medline: 34581679]
70. Chittaro L, Sioni R, Crescentini C, Fabbro F. Mortality salience in virtual reality experiences and its effects on users' attitudes towards risk. Int J Hum Comput Stud 2017 May;101:10-22. [doi: 10.1016/j.ijhcs.2017.01.002]
71. Namkoong K, Chen J, Leach J, et al. Virtual reality for public health: a study on a VR intervention to enhance occupational injury prevention. J Public Health (Bangkok) 2023 Mar 14;45(1):136-144. [doi: 10.1093/pubmed/fdab407]
72. Pashler H, Harris CR. Is the replicability crisis overblown? Three arguments examined. Perspect Psychol Sci 2012 Nov;7(6):531-536. [doi: 10.1177/1745691612463401] [Medline: 26168109]
73. Schmidt S. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. Rev Gen Psychol 2009 Jun;13(2):90-100. [doi: 10.1037/a0015108]
74. Open Science Collaboration. Estimating the reproducibility of psychological science. Science 2015 Aug 28;349(6251). [doi: 10.1126/science.aac4716]
75. Klein SB. What can recent replication failures tell us about the theoretical commitments of psychology? Theory Psychol 2014 Jun;24(3):326-338. [doi: 10.1177/0959354314529616]
76. Ahn S. Virtual exemplars in health promotion campaigns: heightening perceived risk and involvement to reduce soft drink consumption in young adults. J Media Psychol 2018 Apr;30(2):91-103. [doi: 10.1027/1864-1105/a000184]
77. Ahn SJ, Fox J, Dale KR, Avant JA. Framing virtual experiences: effects on environmental efficacy and behavior over time. Comm Res 2015;42(6):839-863. [doi: 10.1177/0093650214534973]
78. Behr KM, Nosper A, Klimmt C, Hartmann T. Some practical considerations of ethical issues in VR research. Pres Teleoperators & Virtual Environ 2005 Dec;14(6):668-676. [doi: 10.1162/105474605775196535]

Abbreviations

HMD: head-mounted display
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews
VE: virtual environment
VR: virtual reality

Edited by L Riedemann, T Leung; submitted 13.06.23; peer-reviewed by B Jacob, B Cieślik, L Mendes; revised version received 27.11.24; accepted 27.11.24; published 20.01.25.

Please cite as:
Bonneterre S, Zerhouni O, Boffo M
Immersive Virtual Reality for Health Promotion and Primary Prevention in Psychology: Scoping Review
JMIR XR Spatial Comput 2025;2:e49923
URL: https://xr.jmir.org/2025/1/e49923
doi:10.2196/49923

Virtual Humans in Virtual Reality Mental Health Research: Systematic Review

Shu Wei[1,2], PhD; Daniel Freeman[3,4], DClinPsy; Aitor Rovira[3,4], PhD

[1]Department of Psychiatry, University of Oxford, Warneford Hospital, Warneford Ln, Headington, Oxford, United Kingdom

[2]Department of Paediatrics, Yale School of Medicine, New Haven, CT, United States

[3]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom

[4]Oxford Health NHS Foundation Trust, Oxford, United Kingdom

Corresponding Author:
Shu Wei, PhD
Department of Psychiatry, University of Oxford, Warneford Hospital, Warneford Ln, Headington, Oxford, United Kingdom

Abstract

Background: Virtual reality (VR) is showing increasing promise for assessing, understanding, and treating mental health difficulties. Virtual humans (VHs) represent a key aspect within many VR mental health applications. While VHs can play diverse roles and display varied characteristics, their design and influence have rarely been the primary focus of mental health research.

Objective: We aimed to carry out a systematic review of how VHs in immersive VR have been used in applications for mental health, focusing on their roles and interaction types, and the human characteristics being tested.

Methods: Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, we searched PubMed, MEDLINE, PsycINFO, Scopus, and Web of Science, using defined keyword combinations involving VR, VHs, and mental health. Eligible studies included peer-reviewed research using immersive VR with VHs in a mental health context, without restrictions on study design or population. We excluded nonimmersive VR, nonmental health applications, and papers without empirical data. Data were synthesized narratively, and a taxonomy to categorize VHs that we developed was used.

Results: A total of 79 studies met all eligibility criteria. VHs were most frequently applied in studies on social anxiety (n=18), eating disorders (n=18), and psychosis (n=15). They were primarily used as active social interaction partners (n=40), as part of virtual crowds (n=16), and as virtual bodies for participants (n=23). Explicit interactions dominated active partner studies, while implicit and passive or no interactions were prevalent in crowd and body studies. Over half of the studies (n=44) varied the VH characteristics, with body size and gender being the most common variables, and personality was explored in fewer studies (n=5). Only a limited number of studies provided detailed descriptions of VH appearance and behavior, with some including still images and videos.

Conclusions: VHs are versatile tools to be used within VR mental health applications, but their design features are inconsistently reported and insufficiently examined in relation to intervention outcomes. Evidence is limited by heterogeneity in study aims, designs, and populations, and by incomplete reporting of VH characteristics, which constrains replication and cross-study comparison. Standardized reporting and systematic investigations of VH design are needed to optimize their roles in evidence-based mental health applications.

Trial Registration: PROSPERO CRD42021244748; https://www.crd.york.ac.uk/PROSPERO/view/CRD42021244748

(JMIR XR Spatial Comput 2025;2:e75087) doi:10.2196/75087

KEYWORDS

virtual reality; virtual human; agent; avatar; mental health

Introduction

Virtual humans (VHs) have been part of virtual reality (VR) mental health applications since the late 1990s. The pioneering work by Riva and Melis [1] used images of a female silhouette in VR to represent the participant's body, marking one of the earliest uses of VHs in mental health research. Other early studies showed that virtual audiences could effectively induce anxiety in participants [2,3]. Since approximately 2013, advancements in VR technology and reductions in cost have significantly improved accessibility, enabling the use of VHs in research areas such as understanding paranoia [4,5], managing autism-centered interventions [6,7], and improving symptoms from posttraumatic stress disorder (PTSD) [8]. Some recent applications have integrated virtual coaches to guide patients and automate treatments [9,10], as well as body swapping

experiences that allow participants to switch between different entities to experience perspective changes [11-13].

VHs enable researchers to simulate social interactions commonly encountered in daily life [14,15] or under special circumstances such as conflict or danger [16,17]. They can also serve as the self-representations of participants through the sense of embodiment and the creation of the illusion that the virtual body they see and control is their own body [18]. Compared to experimental setups with trained real actors, VHs provide greater experimental control and repeatability, preserving ecological validity [19].

Despite their growing use and potential, VHs have not been studied thoroughly in mental health research. This gap is partly due to the inherently multidisciplinary nature of the field, which requires expertise in psychology, immersive technologies, and 3D design to build VHs that are clinically suitable and technically feasible. Furthermore, conducting full-scale clinical trials to evaluate VH design within patient populations is often impractical. In contrast, the computer science community has explored VHs to optimize their visual and animation quality [20-22], develop realistic and adaptive behaviors [23-25], and create open-source character libraries [26-28]. These technical advances offer promising opportunities to extend the applications of VHs in mental health beyond their current use.

Understanding the role and impact of VHs in mental health research is essential for optimizing current implementations and discovering future applications. As VR and character creation software tools improve, VHs have the potential to play an even more important role in mental health, offering customized therapeutic experiences with fine-tuned emotional attributes designed specifically for each patient. We present a systematic review of research studies in mental health VR featuring VHs, from the early development of VR until November 2024. This review proposes a taxonomy based on the most important aspects to consider in VHs for mental health applications and classifies the studies found in this domain. The systematic review addresses the following questions:

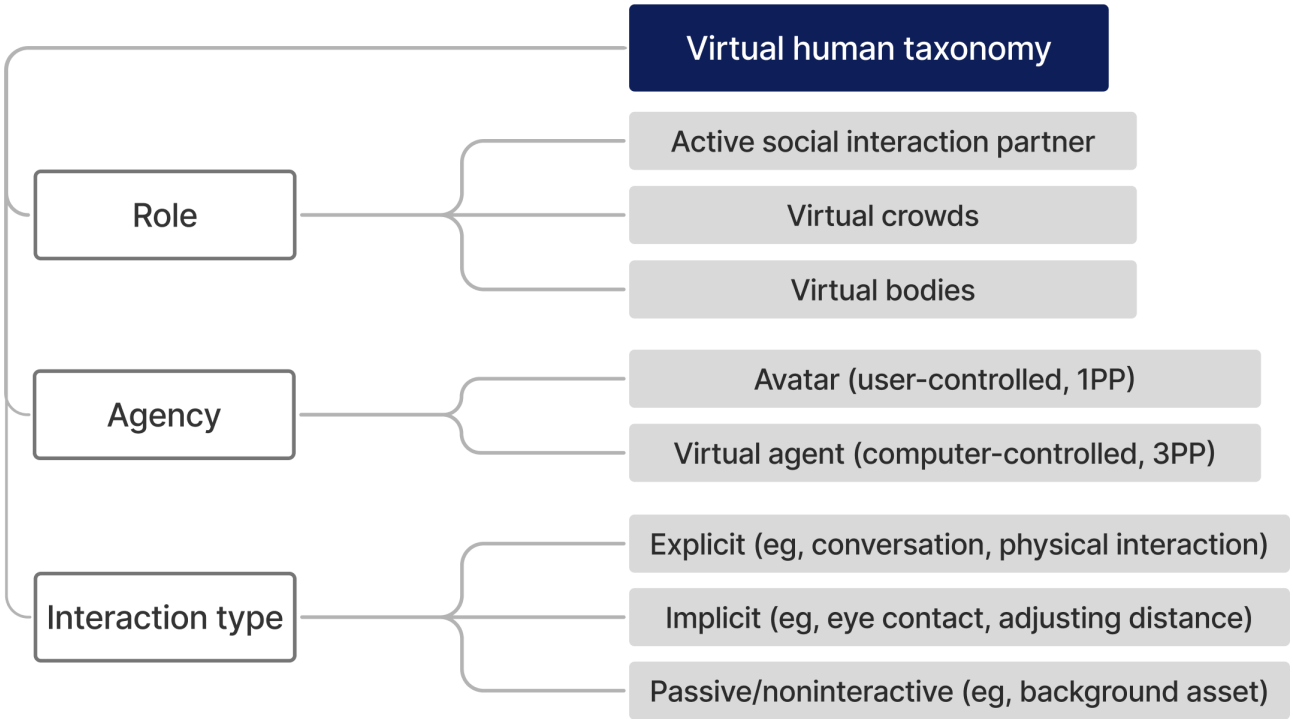
- What roles and functions have VHs played in VR mental health research?
- What have been the chosen agencies for VHs in VR mental health research?
- What characteristics of VHs have been examined in VR mental health research?

Methods

Categorizing VHs in Mental Health VR

We present a taxonomy of VHs in mental health VR (Figure 1), with categorization of their features into 3 main aspects: role, agency, and interaction type. These features were selected because they represent the critical dimensions influencing how VHs are designed, implemented, and experienced in mental health VR. Role defines the primary function of the VH within the virtual environment, agency determines the level of control and autonomy the VH exhibits, and interaction type describes the nature and depth of engagement between the VH and the participant. Together, these features provide a framework for understanding the diverse ways VHs are used in mental health research, enabling researchers to systematically evaluate their design and impact.

Figure 1. A taxonomy of virtual humans based on existing mental health virtual reality studies. 1PP: first-person perspective; 3PP: third-person perspective.



This framework was inductively derived from the included studies, and it was also supported by prior classification approaches in VR and virtual character research. Existing taxonomies highlight comparable dimensions—distinguishing

characters according to their functional role [29,30], level of autonomy or agency [19,29,31,32], and communicative capacity or interaction features [19,30]. By aligning with these established design variables while adapting them to the specific context of mental health VR, our taxonomy provides both empirical grounding and theoretical consistency with prior frameworks.

Role

The role of VHs can be divided into 3 categories: active social interaction partners, virtual crowds, and virtual bodies. These categories were chosen to reflect the distinct ways VHs are used in VR mental health research across diverse applications. They emerged from recurring functional patterns observed across studies and align with prior literature around VHs, which supports the distinction among interactive agents, crowds, and avatars as the most common roles [33-35]. For example, Kyrlitsias and Michael-Grigoriou [33] differentiated VHs by their interaction scenario and level of engagement, while other researchers have examined the psychological impacts of virtual crowds compared to individual characters [34,36].

Active social interaction partners engage with participants to simulate interactions through verbal and nonverbal communication [13,37]. A virtual coach is an example of an active interaction partner that focuses on providing guidance or feedback to facilitate psychological treatment [9]. Virtual crowds, unlike active social interaction partners, have limited contact with participants or normally interact passively with implicit interaction, such as making eye contact [38]. It is common for virtual crowds to be part of the stimuli in the virtual world [39]. Virtual bodies focus on body representation, either by depicting a range of body types [1] or serving as the user's avatar for embodiment experiences [40].

Agency Avatars and Agents

Agency refers to the state or entity of taking action or exerting control [41] and can be categorized as avatars and virtual agents (VAs).

An avatar is a digital representation controlled by a user, serving as their own body in a different medium, for example, in a virtual environment [41]. Typically viewed from a first-person perspective, avatars are co-located with the user's real body. The real body is usually concealed, thus overriding them and creating the illusion that the body they see is their own body. The term "avatar" has often been misused in VR studies to describe any VH, even those who do not represent any user and are controlled by a computer.

A VA is a computer-generated character designed to simulate human interaction autonomously, moving and responding without being controlled by a real person [42]. VAs are mainly controlled by computers and typically viewed from a third-person perspective. They can exhibit varying degrees of autonomy, with semiautonomous agents combining computer automation with operator intervention. This approach is referred to as the Wizard of Oz technique [43] (a hidden operator simulates autonomous system behavior) and allows researchers to make VAs respond in real-time based on participant responses while still being computer-animated [19].

Interaction Types of VAs

VAs can have different degrees of interaction with human participants. We have categorized them into explicit, implicit, and passive or noninteractive.

Explicit interactions involve direct engagement between a VA and a participant, such as conversations, physical interactions (eg, shaking hands [44]), or collaborative activities (eg, playing a sport together [45]). Explicit interactions aim to replicate real-life social experiences, and they can be difficult to implement due to the complexity of managing a large pool of possible outcomes.

Implicit interactions include subtle behaviors, such as acknowledging the participant's presence by making eye contact or adjusting interpersonal distance, without fully engaging in an explicit interaction. Such interactions are less resource-intensive than explicit ones and still contribute to enhancing the sense of presence [46,47], making participants feel as if they are part of the scene rather than just being spectators.

Passive or noninteractive VAs act as background elements in VR environments, unresponsive to participant actions. They are easier to implement due to the scripted nature of their animations. They are used for crowd simulations and creating social situations such as populating a virtual restaurant [48] or constructing an office scene [49].

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for eligible studies are summarized in Textbox 1. In brief, studies were required to (1) use VHs within an immersive VR setting; (2) involve assessment or intervention related to mental health conditions; and (3) include empirical data from human participants. Studies were excluded if they were not related to mental health, did not involve a VR-based VH, or were not available as peer-reviewed full-text articles in English.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Population: Human participants
- Technology use: Immersive virtual reality (VR) with virtual human (VH) characters
- Application area: Mental health assessment or intervention
- Study type: Empirical study involving more than one participant with data collection
- Other criteria: English; available as full text

Exclusion criteria

- Population: Nonhuman studies
- Technology use: VR without VH or nonimmersive VR (eg, desktop system)
- Application area: Not related to mental health
- Study type: Reviews, protocols, abstracts only, commentaries, and opinion pieces
- Other criteria: Non-English; conference abstract

Search Strategy

The literature search was performed in PubMed, Scopus, MEDLINE (Ovid), PsycINFO (Ovid), and Web of Science (Clarivate), with title and abstract searches for keyword combinations involving (1) VR, (2) VH, and (3) mental health. The keyword query was structured as follows: [(virtual reality OR immersive virtual reality OR VR) AND (virtual human OR virtual character OR virtual agent OR avatar OR humanoid) AND ((assessment OR treatment OR therapy OR mental health) OR (mood disorders OR depress* OR bipolar OR mania OR paranoia OR psychosis OR psychotic OR schizophren* OR schizotyp* OR delus* OR hallucinat* OR phobias OR obsessive compulsive disorder OR OCD OR anxiety OR post traumatic stress disorder OR PTSD OR trauma OR anorexia nervosa OR bulimia nervosa OR eating disorders OR binge eating OR insomnia OR sleep OR nightmares OR circadian OR panic OR substance OR abuse OR cannabis OR tobacco OR alcohol OR amphetamine OR hallucinogens OR heroin))]. Details of the searches for each database in the corresponding platforms can be found in Multimedia Appendix 1. A completed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [50] is presented in Checklist 1.

Article Screening

Following database searches, we identified 2596 records (Scopus, n=1232; Web of Science, n=712; PubMed, n=233; APA PsycINFO, n=247; MEDLINE, n=172) between 1997 and 2024 (final search: November 20, 2024). After removing 1085 duplicates (11 manually and 1074 via Covidence), 1511 unique records remained. Title and abstract screening excluded 1280 records that did not involve a VH character, did not use immersive VR, were unrelated to mental health, were nonhuman studies, or were non-English studies. The remaining 231 full-text articles were then assessed in detail, with the reasons for exclusion including the following: not immersive VR (n=73), not related to mental health (n=31), not full paper (eg, conference abstract) (n=21), no empirical data (n=19), no VH (n=11), and case study with only 1 participant (n=3).

Results

Search Results

Figure 2 (PRISMA diagram [50]) summarizes the search and screening results. In total, 79 studies were included in the review: 73 that met all eligibility criteria at full-text screening and 6 identified through citation chaining. We have provided a list of studies that primarily used VHs as active social interaction partners (Table 1), virtual crowds (Table 2), and virtual bodies (Table 3), along with a brief summary for each.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. VR: virtual reality.

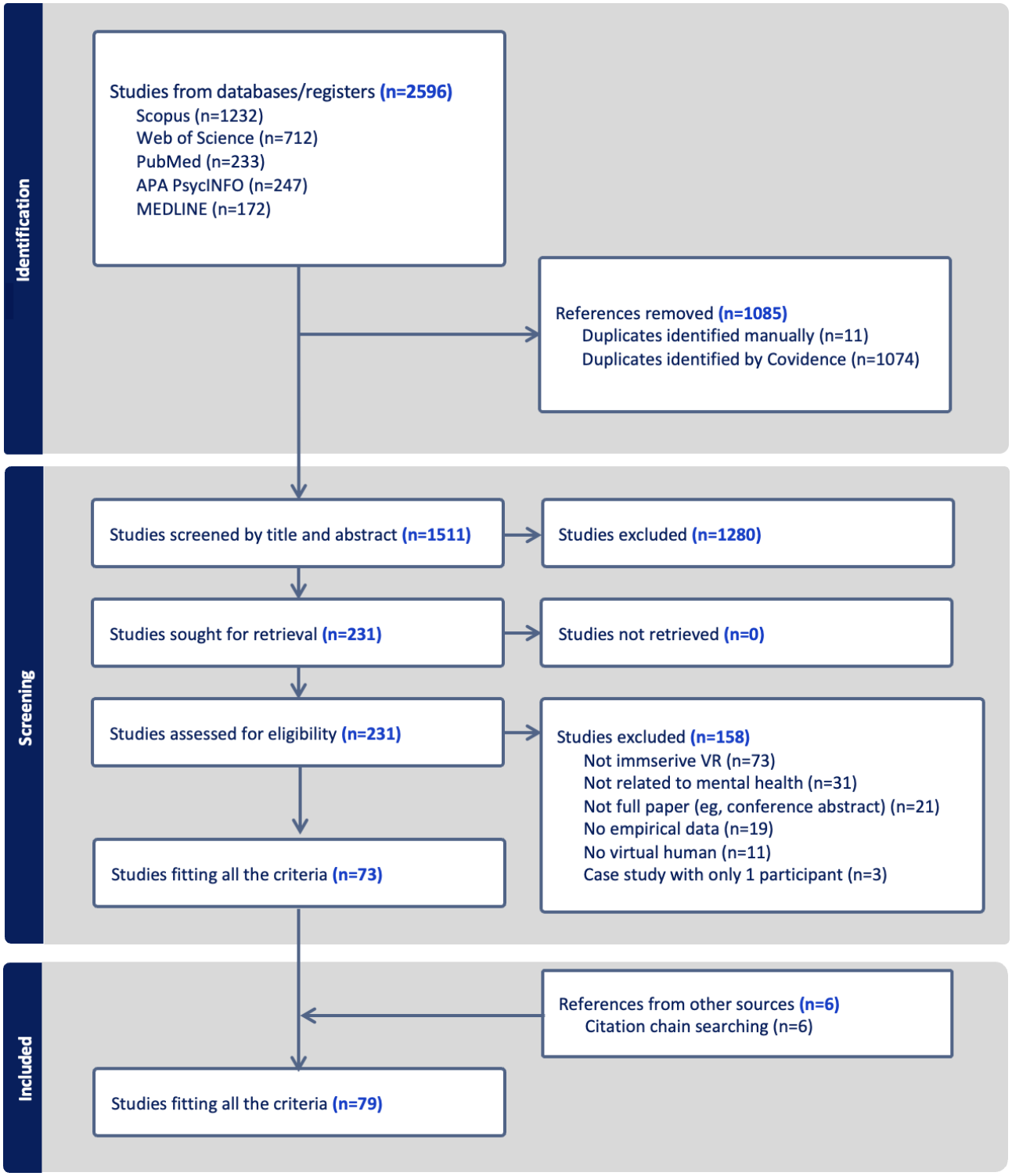


Table . Studies that primarily used VHs[a] as active social interaction partners.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Lee et al [51], 2004	Active interaction partner	Agent	Implicit interaction	—[b]	Substance abuse	16	Males who smoked at least 10 cigarettes a day	VHs in cue exposure therapy exerted a stronger impact compared to seductive objects.
Park et al [52], 2009	Active interaction partner	Agent	Explicit interaction	Gender; Emotional facial expression	Psychosis	54	Patients diagnosed with schizophrenia and the nonclinical general population	VHs with emotional facial expressions successfully triggered emotional and biological responses.
Wieser et al [53], 2010	Active interaction partner	Agent	Implicit interaction	Gender; Gaze pattern; Interpersonal distance	Social anxiety	39	Females with social anxiety	Women paid more attention to the weight-gain areas on VHs than men.
Kwon et al [54], 2013	Active interaction partner	Agent	Explicit interaction	—	Social anxiety	20	Students with social anxiety	VHs provoked both stronger anxiety and presence than those in a nonimmersive display.
Powers et al [55], 2013	Active interaction partner	Semiautonomous agent	Explicit interaction	—	Social anxiety	26	Female undergraduates	Encountering VHs could trigger strong anxiety and fear.
Han et al [56], 2014	Active interaction partner	Agent	Explicit interaction	Attitude (positive, negative dialog)	Psychosis	45	Patients with schizophrenia and the nonclinical general population	Patients with schizophrenia had active avoidance of eye contact with VHs compared to healthy controls.
Hartanto et al [57], 2014	Active interaction partner	Semiautonomous agent	Explicit interaction	Attitude (positive, negative dialog)	Social anxiety	24	General population	Positive dialog feedback from VHs resulted in less anxiety, lower heart rate, and longer answers.
Pan et al [58], 2015	Active interaction partner	Semiautonomous agent	Explicit interaction	Personality (confident, shy)	Social anxiety	24	Male participants	VHs with a shy personality can be perceived as more friendly.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Fornells-Ambrojo et al [4], 2016	Active interaction partner	Semiautonomous agent	Explicit interaction	Responsiveness in social interaction	Psychosis	61	Male participants	Highly contingent VHs were perceived as more trustworthy for extremely paranoid individuals.
Ryan and Griffin [59], 2016	Active interaction partner	Agent	Implicit interaction	—	Social anxiety	27	Students	VR[c] exposure to a VH could successfully trigger social anxiety.
Falconer et al [11], 2016	Active interaction partner	Agent and/or avatar	Explicit interaction	—	Depression	15	Patients with depression	Changing the viewer perspective in a virtual social interaction could decrease self-criticism and increase self-compassion.
Robitaille et al [60], 2017	Active interaction partner	Agent and/or avatar	Explicit interaction	Responsiveness in social interaction	PTSD[d]	12	Nonclinical military members and military members with mild traumatic brain injury	Participants with PTSD displayed a lack of navigational behavior during a VH interaction.
Amaral et al [6], 2018	Active interaction partner	Agent	Implicit interaction	—	Autism	15	Patients with high-functioning ASD[e]	VH affected social attention among patients with autism.
Percie du Sert et al [61], 2018	Active interaction partner	Semiautonomous agent	Explicit interaction	—	Psychosis	15	Patients diagnosed with schizophrenia	VR therapy with VH could improve auditory verbal hallucinations and depressive symptoms.
Freeman et al [62], 2018	Active interaction partner (virtual coach)	Agent	Explicit interaction	—	Phobias	100	Individuals with a fear of heights	A virtual therapist could help automate the VR therapy.
Shin et al [63], 2018	Active interaction partner	Agent	Explicit interaction	Different levels of social pressure	Others	64	General population and patients with internet gaming disorder	Exposure to VHs in gaming cafes triggered gaming cravings.
Kothgassner et al [64], 2019	Active interaction partner	Agent and/or avatar	Explicit interaction	Perceived agency (avatar vs agent)	Social anxiety	56	Students	Virtual social support from a VH could be effective when the recipient thought it was from a human.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Quintana et al [65], 2019	Active interaction partner	Agent	Explicit interaction	—	Social anxiety	53	Female adults	Perception toward the VH could be affected by odorant manipulation.
Reichenberger et al [66], 2019	Active interaction partner	Agent	Explicit interaction	Gender; Interaction pattern (with or without aversive behavior)	Social anxiety	60	General population	VHs could be used to learn about social fear, and women reported higher fear compared to men.
Slater et al [13], 2019	Active interaction partner (virtual coach)	Agent and/or avatar	Explicit interaction	Dialog contents (self-dialog versus scripted dialog)	Depression	58	General population	Self-dialog VR coaching had better results than scripted VR counseling.
Seo et al [67], 2019	Active interaction partner	Agent	Explicit interaction	Interaction pattern (combination of gaze and pointing)	Social anxiety	33	General population	VHs with nonverbal behavior could be used to measure and promote joint attention.
Miloff et al [68], 2020	Active interaction partner (virtual coach)	Agent	Explicit interaction	—	Phobias	70	Patients with a fear of spiders	Alliance toward a virtual therapist is a significant predictor of treatment outcome in VR.
Nijman et al [69], 2020	Active interaction partner	Semiautonomous agent	Explicit interaction	Facial expression	Psychosis	22	Patients with psychotic disorders	VHs' facial expressions can be used to train participants' social cognition.
Kim et al [70], 2020	Active interaction partner	Agent	Explicit interaction	—	Others	36	Nonclinical male volunteers	VHs could be used for assessment and intervention to promote subjective well-being.
Lee et al [71], 2021	Active interaction partner	Agent and/or avatar	Explicit interaction	—	Psychosis	48	Patients diagnosed with schizophrenia and the nonclinical general population	The social context created by VHs impacted the difference in peripersonal space recognition.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Kothgassner et al [45], 2021	Active interaction partner	Agent and/or avatar	Explicit interaction	Perceived agency (avatar vs agent)	Social anxiety	84	Females	Social interaction experience with VHs was comparable to the real-life experience of cyberbullying.
Brander et al [72], 2021	Active interaction partner	Semiautonomous agent	Explicit interaction	Visual and behavioral traits of the virtual character through customization	Psychosis	109	Psychiatric hospital staff (psychotherapists, nursing staff, and administrators)	The function to customize VHs for VR treatment was highly valued by clinical staff.
Guldager et al [73], 2022	Active interaction partner	Agent	Explicit interaction	—	Substance abuse	372	Students aged 15 - 18 years	VR party simulation could improve adolescents' drinking refusal skills.
Freeman et al [9], 2022	Active interaction partner (virtual coach)	Agent	Explicit interaction	—	Psychosis	346	Patients with psychosis	Automated VR therapy with VHs could reduce anxious avoidance and distress in everyday situations.
Fusaro et al [7], 2023	Active interaction partner	Agent and/or avatar	Explicit interaction	—	Autism	53	General population and individuals with ASD	Autistic adults exhibited greater interpersonal distance with VHs.
Bektas et al [74], 2023	Active interaction partner	Agent	Explicit interaction	—	Eating disorder	70	Patients with anorexia nervosa	The addition of a VH to the virtual kitchen scene could increase the association between food-specific state disgust and symptoms of eating disorders.
Giguère et al [75], 2023	Active interaction partner	Semiautonomous agent	Explicit interaction	Voice and facial expressions	Substance abuse	19	Patients with cannabis use disorder and severe mental disorders	The VH intervention could reduce cannabis use among people with cannabis use disorder and severe mental disorders.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Halim et al [12], 2023	Active interaction partner	Agent and/or avatar	Explicit interaction	—	Depression	36	Young adults	Changing the viewer perspective through a virtual body swapping experience can help enhance self-compassion and reduce depressive symptoms.
Wei et al [37], 2023	Active interaction partner (virtual coach)	Agent	Explicit interaction	Head nods and facial expressions	Phobias	120	General population with a fear of heights	The detailed design of a virtual coach's emotional attributes could affect people's therapeutic alliance and their confidence in the VR therapy.
Artiran et al [76], 2024	Active interaction partner	Agent	Explicit interaction	Levels of attention (degrees of backchannels)	Autism	30	Patients with autism and the nonclinical population	The behavior of virtual interviewers influenced the gazes and head movements in individuals with autism.
Natali et al [77], 2024	Active interaction partner	Agent	Explicit interaction	—	Eating disorder	145	Patients with anorexia nervosa	VHs could provide positive social support to patients with anorexia nervosa in a VR food exposure scene.
Hidding et al [78], 2024	Active interaction partner	Agent	Explicit interaction	—	Others	68	Undergraduate students with high levels of self-criticism	Expressing compassion to a VH with similar self-criticism could reduce self-criticism and increase self-compassion.
Freeman et al [10], 2024	Active interaction partner (virtual coach)	Agent	Explicit interaction	—	Psychosis	11	Patients with nonaffective psychosis and low positive self-beliefs (aged 16 - 26 years)	Automated VR therapy with a virtual coach could improve self-beliefs and well-being in young patients with psychosis.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Yamashita and Yamamoto [79], 2024	Active interaction partner (virtual coach)	Agent and/or avatar	Explicit interaction	Identity of the VH (Freud or an intimate person of the participant)	Social anxiety	60	Students with personal problems	VR self-counseling with the intimate other avatar was most effective in improving people's anxiety.
Banakou et al [80], 2024	Active interaction partner	Agent	Explicit interaction	—	Social anxiety	45	General population with public speaking anxiety	Single-session VR exposure therapy with gradually increasing stimulus intensity may be as effective as multisession exposure for public speaking anxiety.

[a]VH: virtual human.

[b]Not applicable.

[c]VR: virtual reality.

[d]PTSD: posttraumatic stress disorder.

[e]ASD: autism spectrum disorder.

Table . Studies that primarily used VHs[a] as virtual crowds.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Slater et al [3], 1999	Virtual crowds	Semiautonomous agent	Implicit interaction	Attentional focus level	Social anxiety	10	General population	A virtual audience could be used to treat public speaking fears; Higher perceived audience interest reduced public speaking anxiety.
Tarnanas et al [49], 2003	Virtual crowds	Agent	Passive interaction	__[b]	Social anxiety	120	General population and patients with work-related stress disorders	Virtual crowds triggered psychological and biological reactions in social anxiety.
Freeman et al [81], 2003	Virtual crowds	Agent	Passive interaction	—	Psychosis	24	General population	People tend to attribute their mental states to VHs.
Freeman et al [82], 2005	Virtual crowds	Agent	Passive interaction	—	Psychosis	30	People covering the whole range of paranoia	Neutral characters could elicit people's persecutory thoughts and be used to understand psychosis.
Freeman et al [15], 2008	Virtual crowds	Agent	Passive interaction	—	Psychosis	200	General population	Neutral characters could elicit paranoid ideation among the general population.
Cho et al [83], 2008	Virtual crowds	Agent	Implicit interaction	—	Substance abuse	10	Population with a low level of alcohol dependence	The presence of a VH in cue exposure induced stronger cravings.
Rizzo et al [39], 2010	Virtual crowds	Agent	Implicit interaction	—	PTSD[c]	20	Active duty service members	A VR[d] environment with VHs could simulate a war scene for the clinical treatment of PTSD and depression.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Brinkman et al [48], 2011	Virtual crowds	Agent	Passive interaction	Density; Ethnic appearance (White European or North African)	Social anxiety	26	General population and patients diagnosed with schizophrenia	VR exposure to an increased density and proportion of VHs with other ethnicities induced strong physiological arousal and distress.
Shiban et al [84], 2015	Virtual crowds	Agent	Passive interaction	—	Social anxiety	40	General population	VHs can be used to understand social fears in VR.
Mountford et al [85], 2016	Virtual crowds	Agent	Passive interaction	—	Eating disorder	18	General population	The exposure to VHs did not necessarily change people's in-state body image.
Atherton et al [86], 2016	Virtual crowds	Agent	Passive interaction	—	Psychosis	26	Males reporting paranoid ideation within the past month	People's own self-confidence could affect their paranoid thoughts toward the VHs.
Gürerk et al [14], 2019	Virtual crowds	Agent	Passive interaction	Different working behaviors	Social anxiety	108	General population	VHs with better working performance could motivate people to perform better.
Takac et al [87], 2019	Virtual crowds	Agent	Implicit interaction	Crowd size; Attentional focus level	Social anxiety	19	General population	Virtual audiences could elicit public speaking distress, which could last beyond VR sessions.
Wei et al [88], 2024	Virtual crowds	Agent	Implicit interaction	Facial expressions	Psychosis	122	General population with elevated paranoia	The detailed design of VH faces affected people's paranoid interpretations.
Gayer-Anderson et al [89], 2025	Virtual crowds	Agent	Implicit interaction	—	Psychosis	481	Adolescents	VR school scenes with VHs could help understand the precursors of adolescents' paranoid ideation.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Rovira et al [90], 2024	Virtual crowds	Agent	Explicit interaction	—	Substance abuse	100	People who smoke daily	VR exposure to smoking-cue VHs elicited stronger cigarette cravings than a neutral environment without VHs.

[a]VH: virtual human.

[b]Not applicable.

[c]PTSD: posttraumatic stress disorder.

[d]VR: virtual reality.

Table . Studies that primarily used VHs[a] as virtual bodies.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Riva and Melis [1], 1997	Virtual body	Agent	Passive interaction	Body size	Eating disorder	119	General population	Virtual figures could be used to understand people's body images.
Perpiñá et al [91], 1999	Virtual body	Agent	Passive interaction	Body size	Eating disorder	13	Patients with eating disorders	VR[b] body image treatment could have a stronger impact than standard treatment.
Aymerich-Franch et al [92], 2014	Virtual body	Avatar	Passive interaction	Different levels of visual similarity to the participant	Social anxiety	187	General population	Having an avatar with an insimilar face could reduce anxiety.
Ferrer-Garcia et al [40], 2017	Virtual body	Agent	Passive interaction	Body size	Eating disorder	23	College students	Owning a fatter virtual body triggered higher levels of body anxiety and fears of weight gain.
Ferrer-Garcia et al [93], 2018	Virtual body	Agent	Passive interaction	Body size	Eating disorder	40	Female students	Owning a virtual body with different body sizes produced changes in body dissatisfaction.
Porras-Garcia et al [94], 2018	Virtual body	Agent	Passive interaction	Body size	Eating disorder	35	College students	Owning a larger-size virtual body increased people's body checking behaviors.
Porras-Garcia et al [95], 2019	Virtual body	Agent	Passive interaction	Body size	Eating disorder	85	College students	Women paid more attention to the weight-gain areas on the virtual figure than men.
Porras-Garcia et al [96], 2019	Virtual body	Agent	Passive interaction	Body size	Eating disorder	50	Undergraduates	Participants embodied through a synchronous visuotactile technique had higher anxiety-associated body dissatisfaction.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Mertens et al [97], 2019	Virtual body	Agent	Passive interaction	—[c]	Phobias	48	General population and individuals with a fear of spiders	Embodying participants into a virtual body from the 1PP[d] could improve the validity of fear conditioning.
Provenzano et al [98], 2020	Virtual body	Agent	Passive interaction	Body size	Eating disorder	20	Patients with eating disorders	Different perspectives (1PP vs 3PP[e]) of viewing the virtual body could change people's body image perception.
Scarpina et al [99], 2019	Virtual body	Agent	Passive interaction	Body size	Eating disorder	30	General population and participants with obesity	A body ownership illusion could be created for both people with obesity and the general population.
Spanlang et al [100], 2018	Virtual body	Agent	Passive interaction	Interaction pattern	Psychosis	27	General population	The body ownership illusion could trigger people's biological reactions.
Fisher et al [18], 2020	Virtual body	Agent	Passive interaction	Body size	Eating disorder	31	Female adolescents	Body size evaluation results from VR were comparable to paper-based measurements.
Porras-Garcia et al [101], 2020	Virtual body	Agent	Passive interaction	Body size	Eating disorder	41	Female college students	People could have an attentional bias toward the embodied virtual figure because of their own body dissatisfaction.
Gorisse et al [102], 2021	Virtual body	Agent and/or avatar	Explicit interaction	Interaction pattern (engagement level with virtual crowds)	Psychosis	30	General population	Observing a body double successfully engage in a VR social interaction could reduce paranoia.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Wolf et al [103], 2021	Virtual body	Agent and/or avatar	Passive interaction	—	Eating disorder	56	Females with normal weight	Embodiment through a VH could affect people's self-evaluation of their body weight.
van Gelder et al [104], 2022	Virtual body	Agent and/or avatar	Explicit interaction	Age (current self and future self)	Others	24	Convicted male offenders	Offenders' interactions with a VH representing their future self could reduce their self-defeating behaviors.
Yamamoto and Nakao [105], 2022	Virtual body	Agent	Passive interaction	Color	Depersonalization	31	Males with various depersonalization tendencies	People with depersonalization had weaker body ownership of the self-representative figure.
Burin et al [106], 2022	Virtual body	Agent and/or avatar	Passive interaction	Embodiment perspectives (1PP or 3PP)	Others	52	General population	Experiencing a moving virtual body from the 1PP could enhance stress coping ability.
Vahle and Tomasik [107], 2022	Virtual body	Avatar	—	Age (age-congruent and age-incongruent with the participants)	Others	74	General population aged between 18 and 30 years	The embodiment of an older avatar increased young adults' social motivation.
Mendoza-Medialdea et al [108], 2023	Virtual body	Avatar	—	Body size	Eating disorder	57	College women	Manipulating the VH body size could produce changes in body dissatisfaction.
Ascione et al [109], 2023	Virtual body	Avatar	—	—	Eating disorder	23	Adolescent patients with anorexia nervosa	The VR body attentional bias modification task could reduce body-related attentional bias among people with anorexia nervosa.

Study	Role	Agency	Virtual agent interaction type	Parameters for manipulation	Condition	Participants, n	Participant characteristics	Comments on findings
Schroeder et al [110], 2023	Virtual body	Avatar	—	Body size	Eating disorder	43	Women with high or low body dissatisfaction	People with low body dissatisfaction showed stronger disembodiment during self-avatar mirror exposure.

[a]VH: virtual human.

[b]VR: virtual reality.

[c]Not applicable.

[d]1PP: first-person perspective.

[e]3PP: third-person perspective.

Mental Health Conditions

VHs were used in the following research areas in mental health: social anxiety (n=20), eating disorders (n=18), psychosis (n=17), substance abuse (n=5), phobias (n=4), depression (n=3), autism (n=3), PTSD (n=2), dissociative disorders (n=1), and others without specific mental health conditions (n=6).

Roles of VHs

We categorized the studies based on our taxonomy. Studies with overlapping roles were assigned to their primary role or the one with the highest interactivity. The roles of VHs in the included studies could be categorized into active social interaction partners (n=40), virtual crowds (n=16), and virtual bodies (n=23).

Active Social Interaction Partners

In 40 studies, VHs primarily served as active social interaction partners, tested in three different ways: (1) manipulating specific VH characteristics to assess their influence, (2) comparing the responses of clinical and nonclinical individuals, and (3) validating whether VHs in VR could trigger participants' responses or replicate the effects of real-world experiences.

Manipulating VH Characteristics

Twelve studies explicitly examined VH characteristics in social interactions. Wieser et al [53] investigated the anxiety-inducing effects of VH gender and gaze behaviors, finding that male agents with direct gaze elicited stronger anxiety and heart rate elevation among socially anxious women compared to female agents. A similar gender-related effect was observed in a VR fear conditioning program by Reichenberger et al [66], where female patients reported higher anxiety to aversive behaviors (eg, unpleasant air blasts) from male VAs than from female VAs. Park et al [52] and Nijman et al [69] focused on nonverbal behaviors, implementing different facial expressions on VHs to help individuals with psychotic disorders practice emotion recognition.

Researchers also manipulated the dialog responses (positive or negative) [57] and responsive frequency (high or low) [4] in conversation agents. Both the attitude and reply frequency of VHs in a conversation affected participants' behavioral (eg, interpersonal distance) and physiological (eg, heart rate) responses. In another study by Slater et al [13], participants engaged in a self-conversation by switching between 2 virtual bodies: first embodying a VH resembling themselves to describe their problem to a virtual Sigmund Freud and then embodying Freud to offer advice to their virtual self. This method resulted in a better perception of help compared to scripted dialog. A similar design was then adopted by Yamashita and Yamamoto [79], who found that VR self-counseling with an intimate other avatar was most effective in reducing anxiety. Furthermore, Pan et al [58] explored the impact of VH personality traits (confident vs shy) on participants' perceptions, showing that the shy VH was perceived as more positive and friendly. The effects of VH emotional attributes were further demonstrated by Wei et al [37], who showed that a virtual coach with positive facial expressions and affirmative nods enhanced participants' therapeutic alliance and treatment confidence during a VR fear-of-heights consultation. Lastly, the perceived agency of VHs was studied by Kothgassner et al [45,64], who compared participants' reactions when they believed they were interacting with a human versus a computer. Both experiments indicated that interactions with VHs can be as effective as interactions with real-life counterparts, provided participants believe the VH is controlled by a human in real time.

Comparing Clinical and Nonclinical Populations

In another use case, 5 studies used active interaction partners to assess the differences between clinical and nonclinical populations. Han et al [56] found that patients with schizophrenia avoided eye contact more often when interacting with 2 VHs in a 3-person conversation. Lee et al [71] used a ball game to study self-disturbances (eg, a distorted perception of one's own body boundary), suggesting that patients with schizophrenia showed longer reaction times and a weaker sense of body boundary compared to controls. Robitaille et al [60] designed a scenario to detect executive functioning difficulties in military personnel with traumatic brain injury. In a military patrol scene, a virtual male dressed as a civilian was used to simulate a divided attention task by approaching the participant. Both groups tolerated the experience, but those with posttraumatic brain injury displayed navigation difficulties after a divided attention task was added. Lastly, Fusaro et al [7] and

Artiran et al [76] studied the differences in responses to VHs between participants with autism and nonclinical controls, and found that individuals with autism maintained greater interpersonal distance and responded less to VHs' social cues (eg, head turning and direct gaze) during virtual interviews.

Validation Studies With VHs

Twenty-two studies used active interaction partners to understand whether VHs in VR could trigger participants' responses or replicate the effects of a real-world experience. Ten studies involved participants with mental health conditions or specific phobias. Three of them were related to psychosis. Percie du Sert et al [61] tested a therapy for auditory verbal hallucinations in psychosis. Patients interacted with a VH representing the harmful voice they heard, and this exposure reduced their auditory verbal hallucination–related depressive symptoms. Freeman et al [9] developed gameChange, an automated VR therapy designed to address agoraphobia in psychosis. Patients were guided by a virtual coach through everyday scenarios like visiting a general practitioner or shopping, resulting in significant anxiety reduction, particularly in those with severe symptoms. More recently, Freeman et al [10] launched Phoenix, where a coach named Farah helps young patients with psychosis build self-confidence through guided tasks. Two other studies focused on eating disorders. Bektas et al [74] used a virtual kitchen scene in patients with anorexia nervosa, finding that a VH encouraging patients to challenge their disorder elicited stronger disgust reactions toward food. Natali et al [77] used the same scenario to induce positive mood, showing that supportive VHs can reduce anxiety and improve mood in patients with anorexia. Regarding phobias, Freeman et al [62] conducted a clinical trial for treating fear of heights. A VR coach called Nic guided participants through the different stages of the experience, leading to a significant reduction in their fear. Similarly, Miloff et al [68] used a holographic coach for spider phobia, where positive relationships predicted better outcomes. Falconer et al [11] promoted self-compassion in depression through role-swapping between a compassionate adult and a virtual child. Amaral et al [6] conducted a clinical trial using VHs for social and cognitive training in individuals with autism. A further study [75] applied VR to treat addiction to cannabis. Participants interacted with a VH representing a key figure in their cannabis consumption, such as a friend, a drug dealer, or those triggering their cravings, showing potential to reduce usage.

Among the remaining 13 validation studies involving nonclinical populations, 6 focused on social anxiety. Kwon et al [54] compared the effects of displaying VHs on a desktop monitor versus an immersive VR system, finding that immersive VR elicited stronger anxiety. This was further supported by a VR self-compassion induction [59] and a joint attention study [67], both of which concluded that VR had a greater capacity to induce anxiety compared to nonimmersive setups. Powers et al [55] examined participants' fear levels during VR and real-life interactions, noting that VR interactions triggered fear and anxiety similar to real-life situations. Banakou et al [80] suggested that single-session VR exposure could match multisession efficacy for public speaking anxiety. Moreover, Quintana et al [65] investigated the effects of smells on the perceptions of VHs. They discovered that exposure to fear-related odorants increased anxiety and reduced trust toward the virtual character. Other applications included improving depression via compassion practice [12], testing a customizable VH-based auditory hallucination therapy tool with hospital staff [72], and simulating alcohol refusal scenarios with VHs in VR FestLab [73]. Similar refusal training targeted gaming addiction [63] and nicotine addiction [51]. Finally, VR was used to promote well-being and self-confidence in youth by fostering strength awareness [70] and reducing self-criticism [78] through structured conversations with virtual characters.

Virtual Crowds

Sixteen studies used virtual crowds. Six of them were related to anxiety. In 1 of the earliest studies, Slater et al [3] compared public speaking to an attentive versus a disruptive virtual audience, finding that crowds could evoke anxiety comparable to real life, with better performance when the audience was attentive. Similar findings were reported by Takac et al [87], who suggested that the distress created in a virtual public speaking scenario could have lingering effects after the VR session. Virtual crowds were also used to populate a virtual office, finding that watching other VHs working was effective in triggering work-related stress [49] or creating peer effects to make people work more efficiently [14]. Brinkman et al [48] explored the effect of the size and ethnicity of virtual crowds in a bar scene, showing that VR exposure to an increased density and proportion of VHs with other ethnicities could induce stronger distress. Additionally, Shiban et al [84] incorporated air blasts and the sound of a female character screaming when participants approached a virtual crowd to study how individuals develop social fears and how these fears can be diminished or overcome through repeated exposure.

Six studies used a virtual crowd for research on psychosis. Early tests examined people's appraisals of neutral characters in a library scene, demonstrating that virtual crowds elicited persecutory thoughts for insights into psychosis [81,82]. Later research employed daily VR scenarios populated with VHs, such as in an underground train or school canteen, to understand factors associated with paranoid thinking. The findings suggested that paranoia could be predicted by anxiety, worry, perceptual anomalies, and cognitive inflexibility [15], and was influenced by individuals' self-confidence [86]. For adolescents, interpersonal threat or hostility heightened state paranoia [89]. Further, Wei et al [88,111] found that VH facial design influenced both paranoia and visual attention.

Mountford et al [85] used a virtual crowd in a virtual bus ride to study eating disorders. Participants were tested for their body image satisfaction, and the results showed that exposure to the virtual scenario with VHs did not necessarily change people's current feelings or perceptions about their body image. Cho et al [83] designed a VR bar to study alcohol cravings and found that the introduction of drinking characters led to a stronger drinking desire compared to simply displaying the alcohol in the scene. Similarly, Rovira et al [90] found that VR exposure to smoking-cue VHs elicited stronger cigarette cravings than neutral environments without VHs. Rizzo et al [39] used virtual

soldiers and civilians for military training and found a reduction in PTSD severity in 80% of the study completers.

Virtual Bodies

VHs were used as body representations in 20 studies, with 15 focusing on eating disorders. Riva and Melis [1] conducted the first study in the literature on body image representation using VR. Participants selected virtual figures representing their current and ideal body size. Perpiñá et al [91] introduced a feature allowing patients with eating disorders to adjust the size of a virtual body to reflect their body image. Both showed VR's feasibility for assessing body image disturbances, with results comparable to paper-based measures but showing greater engagement [18]. While these early studies did not induce body ownership, later studies often aimed to do so.

Studies employing the body ownership illusion paradigm found that embodying a larger virtual figure led participants to check their virtual body more frequently [94,95] and experience a higher degree of self-dissatisfaction and anxiety [40,96,98,101,108]. In contrast, Schroeder et al [110] found that body dissatisfaction was not necessarily affected by the weight of the avatar, although people with higher body dissatisfaction exhibited more body-checking behaviors in weight-related areas.

Ferrer-Garcia et al [93] created a paradigm to reduce anxiety related to body image in female students by making them return to the normal-size virtual figure after experiencing a larger body. A similar method was used to alleviate body image distortion in both healthy-weight participants and those with obesity [99]. Wolf et al [103] compared the body weight perceptions of 2 groups of females: one that embodied a VH and another that observed the VH as another person's avatar. The self-embodiment group underestimated the body weight of the VH compared to those who rated the VH as another's avatar. Furthermore, Ascione et al [109] conducted a body-related attention bias modification task to reduce excessive visual focus on the weight-related areas of a self-represented VR body, which significantly decreased body dissatisfaction.

Three other studies used virtual bodies for psychosis research. Spanlang et al [100] examined whether self-fragmentation in VR, where participants embodied a VH while retaining physical presence in the real world, impacted the biological responses in patients with schizophrenia. Yamamoto and Nakao [105] found that perceiving a fake body as one's own reduced body ownership in depersonalization. Gorisse et al [102] used a virtual body double resembling each participant for a social VR task. They concluded that observing a body double engaging in VR social interactions could reduce persecutory thoughts during the task.

Additionally, van Gelder et al [104] demonstrated that alternating convicted offenders between their current and aged future selves reduced self-defeating behaviors. Similarly, embodying an older body could increase social motivation among young adults [107]. Besides, Burin et al [106] presented participants with a moving character from first-person and third-person perspectives. They concluded that a body illusion from the first-person perspective created stronger physiological activation to help people practice stress coping. Aymerich-Franch et al [92] assessed the impact of the facial similarity of a VH on participants, finding that embodying a figure with a dissimilar face reduced anxiety during VR presentations.

VAs and Avatars

Out of the 79 studies, 74 involved the use of VAs. Nine of them adopted a semiautonomous agent design, where facilitators manually triggered an audio effect [3,61], adjusted VH voices [75], or played prerecorded animations [4,55] to customize characters' behaviors in real time.

Regarding the interaction types of VAs, approximately half (39/74, 53%) engaged in explicit interactions with participants, primarily serving as active interaction partners. In 10 studies, VAs exhibited implicit interactions (eg, occasionally looking at participants and then looking away). Passive interactions were observed in 26 studies where VHs were used as virtual crowds and virtual bodies.

In the reviewed studies, the term "avatar" was often used to describe any type of VH, even when they did not represent the participant's self. Only 5 studies used VHs as participant avatars. For example, Aymerich-Franch et al [92] used an avatar with varying levels of visual similarity to the participant's face to study the self-embodiment experience. Additionally, 12 studies combined the use of agents and avatars to create perspective changes through body-swapping experiences in social interactions [11,13,104] or to explore the effects of vicarious agency (the perception of control over the VH) [102]. Kothgassner et al [45,64] further examined the effect of perceived agency by comparing participants' VR experiences with VAs and avatars. The results indicated that avatars could elicit stronger emotional responses compared to VAs.

Human Characteristics

Of the 79 studies reviewed, 44 examined the effects of manipulating specific characteristics of VHs. These manipulations aimed to assess how physical and behavioral attributes of VHs influence participants' perceptions and behavioral responses. The manipulated features included their visual appearance [48,52,99] (eg, gender, ethnicity, and body size), interaction behavior [7,76,87] (eg, interpersonal distance, eye gaze behavior patterns, and responsiveness toward participants), perceived agency [45,64] (VA or avatar), and emotions [52,75] (eg, emotional facial expressions) or personality [58]. While most of the traits were nonverbal behaviors, 4 studies explored the effects of verbal behaviors, including dialog content [13], conversation attitude (eg, positive vs neutral dialog) [56,57], and the voice of VHs [75]. The most frequently studied trait was body size (n=13), while traits, such as sense of affirmation [37] and personality [58], received limited investigation.

In addition to individual traits, 2 studies examined the effects of crowd size and density as independent variables in virtual crowd settings [48,87]. These studies explored how variations in the number and arrangement of virtual characters influence participants' sense of presence and anxiety in crowded environments.

Discussion

Summary of Findings

This paper reviewed VR studies that included VHs in scenarios to understand and treat mental health difficulties. We examined 79 studies and categorized them based on the primary role of the VHs: active social interaction partners, virtual crowds, and virtual bodies. We focused on the use of agency, types of interactions, and characteristics of the VHs. VHs have served diverse purposes and demonstrated the ability to create psychological, behavioral, and physiological influences. However, few studies have provided comprehensive descriptions of VHs' visual appearance or behavioral capabilities. Without such details, it becomes challenging to assess their actual impact, and replicability is reduced. The potential of each VH role requires further exploration to address these gaps and the associated challenges.

Active Social Interaction Partners

Active social interaction partners interact explicitly with participants. A persistent challenge in this context has been enhancing the realism and responsiveness of these interactions, particularly during close-up engagements [19,112,113]. Improving both visual fidelity and behavioral plausibility is essential for achieving lifelike interactions [30,114,115]. Consistent design of both visual and behavioral elements is needed to avoid the uncanny valley, where characters appear almost human but evoke discomfort due to subtle imperfections, as also seen in VR interactions [116,117]. Additionally, it is important to consider VH behavioral plausibility and consistency across multiple behavioral channels, including both verbal and nonverbal communication (eg, body movement, facial expressions, voice, and dialog content). For example, Choudhary et al [118] demonstrated that mismatches between facial expressions and vocal tone (eg, a happy face with an unhappy voice) negatively impacted participants' immersion and engagement, highlighting the importance of aligning these communication modes.

Regarding responsiveness and interactivity, current feedback in mental health VR often relies on prerecorded actions triggered by participant inputs, such as user interface selection, position, or response time [49,60,65]. Future applications could integrate machine learning algorithms, such as reinforcement learning, to create autonomous VHs that learn from experience [119]. These VHs could adjust parameters, such as the interpersonal distance, to maximize attention and keep participants engaged. Algorithms could detect interaction dynamics (eg, participant attitudes and responsiveness) and adjust VH behaviors in real-time to achieve specific therapeutic goals, particularly in applications encouraging positive behaviors. Another area to explore is the real-time monitoring of physiological responses to create adaptive interactions based on participants' affective state. A few of the studies reviewed used biometric data (eg, salivary alpha-amylase, heart rate, and galvanic skin) to evaluate physiological responses [52,100,106], but none leveraged these data to enhance interaction fidelity. Real-time estimates of emotional arousal and valence could enable VHs to respond meaningfully to participants' emotional changes [120].

A virtual coach represents a specific type of active social interaction partner, designed to guide and motivate users toward positive behavioral changes while facilitating the automation of therapy delivery. Focus group feedback indicated that users valued the convenience, accessibility, and anonymity provided by a virtual coach, but they also expressed concerns about the coach feeling unrelatable and difficult to engage with [121]. The development of a VR coach should incorporate knowledge and feedback from domain experts and targeted user groups to ensure it is tailored to the specific age range and mental health conditions it aims to address [122,123]. Generally, increasing the engagement and friendliness of communication from virtual coaches is likely to improve their effectiveness and the outcomes of VR therapy [68].

Virtual Crowds

In virtual crowd studies, it is standard practice to use identical designs and behaviors for characters that share common objectives to achieve a balance between realism and computational efficiency [124,125]. However, visual and behavioral repetition patterns in a group of VHs are very noticeable and decrease fidelity. Introducing variations in appearance and movement is important to diversify crowds and avoid these patterns [126,127]. Visual diversity can be achieved, for example, by randomizing height, body shape, hair color, and clothing. Behavioral diversity can be enhanced by blending animations, using motion capture data from multiple individuals, and implementing procedural animations with a certain degree of randomization to generate varied movements in real-time based on environmental and interaction dynamics.

There are additional considerations. When virtual crowds have autonomy to move around a space, each one should navigate naturally, avoiding collisions with static and dynamic obstacles. For example, Trivedi and Mousas [128] implemented a crowd behavior system for a VR street scene, where each VA had navigation scripts with steering and pathfinding capabilities to avoid collisions. Each agent also had a customizable target location, walking speed, and animation (eg, head direction and gaze pattern). Other advanced virtual crowd control systems generate real-time behaviors driven by user input and the dynamics of other virtual entities within the crowd [129]. These designs could be applied in mental health applications to improve crowd interactions and better understand responses to busy environments.

Virtual Bodies

Studies involving virtual bodies predominantly helped individuals reassess their body image biases or gain new perspectives, provided by the body ownership illusion. These use cases require improved methods to enable the sense of embodiment. Existing methods like visual-motor synchrony [97] and visual-tactile synchrony [96,99] are designed to align visual feedback with participants' movements or tactile sensations to enhance the illusion of ownership over the virtual body. While effective in initiating this illusion, sustaining or manipulating it throughout the VR experience remains challenging. Mismatched sensory feedback or slight discrepancies in motion can disrupt the sense of embodiment [130,131]. Integrating other sensory modalities, such as auditory

feedback (eg, footsteps and voice modulation) or olfactory feedback (eg, simulating a scent experience based on the participant's location), could enhance the body ownership illusion. Combined multisensory feedback has been shown to strengthen and prolong the influence on embodiment, affect, and behaviors [132,133].

Customizing virtual bodies to closely resemble participants can enhance the sense of embodiment. Aymerich-Franch et al [92] found that embodying a virtual body with a similar face increased self-association and anxiety during public speaking tasks compared to dissimilar faces. This provides evidence for the importance of avatar personalization and its potential to modulate emotional responses. Customization should consider not only physical appearance but also factors like ethnic background [134] and voice [135]. These factors can significantly contribute to the body ownership illusion.

Ethical Considerations and Recommendations

The integration of VHs in mental health VR raises potential ethical and methodological challenges. A key concern is the lack of standardization in the technical implementation of character creation and animation. Many implementations rely on proprietary or nonstandardized software, which complicates replication and cross-study comparison. When such tools become inaccessible or difficult to adopt, scaling the experiments or building cumulative evidence becomes challenging. Moving toward open-source platforms or developing shared technical standards would enhance reproducibility and research efficiency. Encouragingly, the computer science community has started to introduce open-source 3D character libraries [26-28], some of which have already been used in mental health research [11,13]. This provides early examples of community-driven resources that could serve as a foundation for broader standardization.

Equally important is the transparency of reporting with regard to VHs. Many current studies provide limited descriptions of VH characteristics, and some even omit a general description of VR scenes. Such omissions reduce study transparency and limit replicability for continuous research. General ethical guidelines from the UK National Institute for Health and Care Excellence (NICE) [136] emphasize clear reporting of system design and capabilities for therapeutic applications. The CONSORT-EHEALTH extension recommends that digital health interventions specify intervention components, delivery modes, tailoring, and technical features [137]. More specifically, RATE-XR (reporting for the early-phase clinical evaluation of applications using extended reality) provides a reporting framework for early-phase clinical extended reality applications, with a checklist covering application characteristics, hardware, software, and development details [138]. For more tailored standards, VH-specific reporting practices, ideally supported by visual and technical documentation (eg, screenshots, videos, and model specifications), are needed to improve reproducibility and comparability across disciplines.

Moreover, data privacy and participant safeguards also require careful consideration. VH-mediated applications can sometimes collect sensitive physiological, behavioral, and conversational data [13,48,139], raising risks related to identity inference and emotional profiling, where user behavior data can sometimes be used to infer sensitive personal traits [140]. In addition, embodiment-based interactions can sometimes provoke unintended psychological effects, such as distress or overidentification with the VH [141,142]. To mitigate these risks, researchers should implement robust informed consent procedures and ensure secure data handling. Safeguards, such as participant prescreening, clinician oversight, built-in safety controls (eg, pause or exit functions in VR software), and structured debriefing protocols, are essential to protect participants.

Limitations

This systematic review has several limitations. First, our search strategy may not have captured all relevant VR mental health studies involving VHs. We used 5 databases and conducted searches limited to titles and abstracts with fixed keyword combinations, excluding studies that did not simultaneously mention VR, VHs, and mental health. To maintain a consistent search strategy across platforms, we did not use controlled vocabulary, such as Medical Subject Headings (MeSH) in PubMed, which may have further reduced recall. To mitigate this, we conducted citation chaining and identified 6 additional eligible studies not captured in the original search. However, the omission may have still remained, and it may have narrowed the evidence base. Future updates should refine the search strategy by incorporating broader synonyms and controlled vocabulary, and including gray literature sources.

Second, there are limitations in our categorization of VH roles and characteristics. Existing taxonomies used to classify virtual characters and their behaviors [30,143] are at least a decade old and do not specifically focus on VR studies in mental health. We developed our own classification based on the usage of VHs in the reviewed studies. This framework is tailored to mental health research, and it may require adaptation for application in other research areas. Ongoing efforts to develop and validate updated, domain-general taxonomies would improve consistency and comparability across studies.

Third, most included studies were pilot, feasibility, or exploratory in nature, often with small sample sizes and limited follow-up. While we included only empirical studies with more than one participant to enhance rigor, the predominance of early-stage designs may have constrained overall study quality and generalizability. Future research could prioritize more robust designs, such as randomized controlled trials or longitudinal studies with quality assessments, to generate stronger and more interpretable evidence.

Fourth, regarding result analysis, we opted for a narrative review approach instead of a quantitative evaluation to measure the quality and effectiveness of VHs across studies. This choice was made due to the variability in study objectives (ranging from validation and assessment to clinical treatments) and the diversity in participant characteristics. In addition, many studies lacked sufficient descriptions of VH attributes and consistent reporting on VR systems and software, which constrained systematic comparison and aggregation. As the field progresses and reporting becomes more standardized, meta-analyses and formal quality appraisals will become feasible.

Conclusions

VHs can be used in many different ways within immersive VR mental health research and have served primarily as active interaction partners, virtual crowds, and virtual bodies. Across the 79 studies reviewed, most VHs were computer-controlled agents, with perceived agency shaping emotional and behavioral responses. Research to date has predominantly examined body size and gender, with limited attention to emotional expression or personality traits. Many studies provided insufficient technical and visual details on VH design, limiting reproducibility and cross-study comparison. Addressing these gaps is crucial for advancing the evidence base and understanding how specific VH characteristics influence mental health outcomes in VR. To our knowledge, this review provides the first synthesis of the use of VHs in VR mental health, offering a potential foundation for future VR programming of VHs and future studies on their implementation.

Acknowledgments

We would like to thank all researchers whose work contributed to the studies reviewed in this paper. This research has been funded through the Oxford Cognitive Approaches to Psychosis (OCAP) research fund.

ChatGPT 4.0 (OpenAI) was used to proofread some original paragraphs, with selected suggestions adopted after careful review.

Authors' Contributions

Conceptualization: SW (lead), AR (equal), DF (equal)
Data curation: SW
Formal analysis: SW
Methodology: SW (lead), AR (supporting), DF (supporting)
Supervision: AR (equal), DF (equal)
Writing – original draft: SW
Writing – review & editing: SW (lead), AR (supporting), DF (supporting)

Conflicts of Interest

DF holds options in XRHealth, a virtual reality mental health treatment company. The other authors do not have any competing interests.

Multimedia Appendix 1
Search query details.
[DOCX File, 21 KB - xr_v2i1e75087_app1.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 277 KB - xr_v2i1e75087_app2.docx]

References

1. Riva G, Melis L. Virtual reality for the treatment of body image disturbances. Stud Health Technol Inform 1997;44:95-111. [Medline: 10175346]
2. Pertaub DP, Slater M, Barker C. An experiment on public speaking anxiety in response to three different types of virtual audience. Presence: Teleoperators & Virtual Environments 2002 Feb;11(1):68-78. [doi: 10.1162/105474602317343668]
3. Slater M, Pertaub DP, Steed A. Public speaking in virtual reality: facing an audience of avatars. IEEE Comput Grap Appl 1999;19(2):6-9. [doi: 10.1109/38.749116]
4. Fornells-Ambrojo M, Elenbaas M, Barker C, et al. Hypersensitivity to contingent behavior in paranoia: a new virtual reality paradigm. J Nerv Ment Dis 2016 Feb;204(2):148-152. [doi: 10.1097/NMD.0000000000000414] [Medline: 26825264]
5. Valmaggia LR, Day F, Rus-Calafell M. Using virtual reality to investigate psychological processes and mechanisms associated with the onset and maintenance of psychosis: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2016 Jul;51(7):921-936. [doi: 10.1007/s00127-016-1245-0] [Medline: 27262562]
6. Amaral C, Mouga S, Simões M, et al. A feasibility clinical trial to improve social attention in autistic spectrum disorder (ASD) using a brain computer interface. Front Neurosci 2018;12:477. [doi: 10.3389/fnins.2018.00477] [Medline: 30061811]
7. Fusaro M, Fanti V, Chakrabarti B. Greater interpersonal distance in adults with autism. Autism Res 2023 Oct;16(10):2002-2007. [doi: 10.1002/aur.3013] [Medline: 37658641]
8. Shilling R. Clinical Virtual Reality tools to advance the prevention, assessment, and treatment of PTSD. Eur J Psychotraumatol 2017;8(sup5):1414560. [doi: 10.1080/20008198.2017.1414560] [Medline: 29372007]
9. Freeman D, Lambe S, Kabir T, et al. Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with

mediation and moderation analyses. Lancet Psychiatry 2022 May;9(5):375-388. [doi: 10.1016/S2215-0366(22)00060-8] [Medline: 35395204]
10. Freeman D, Freeman J, Ahmed M, et al. Automated VR therapy for improving positive self-beliefs and psychological well-being in young patients with psychosis: a proof of concept evaluation of Phoenix VR self-confidence therapy. Behav Cogn Psychother 2024 May;52(3):277-287. [doi: 10.1017/S1352465823000553] [Medline: 37942541]
11. Falconer CJ, Rovira A, King JA, et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open 2016 Jan;2(1):74-80. [doi: 10.1192/bjpo.bp.115.002147] [Medline: 27703757]
12. Halim I, Stemmet L, Hach S, et al. Individualized virtual reality for increasing self-compassion: evaluation study. JMIR Ment Health 2023 Oct 2;10(1):e47617. [doi: 10.2196/47617] [Medline: 37782537]
13. Slater M, Neyret S, Johnston T, et al. An experimental study of a virtual reality counselling paradigm using embodied self-dialogue. Sci Rep 2019 Jul 29;9(1):10903. [doi: 10.1038/s41598-019-46877-3] [Medline: 31358846]
14. Gürerk Ö, Bönsch A, Kittsteiner T, Staffeldt A. Virtual humans as co-workers: a novel methodology to study peer effects. J Behav Exp Econ 2019 Feb;78:17-29. [doi: 10.1016/j.socec.2018.11.003]
15. Freeman D, Pugh K, Antley A, et al. Virtual reality study of paranoid thinking in the general population. Br J Psychiatry 2008 Apr;192(4):258-263. [doi: 10.1192/bjp.bp.107.044677] [Medline: 18378984]
16. Kane HS, McCall C, Collins NL, Blascovich J. Mere presence is not enough: responsive support in a virtual world. J Exp Soc Psychol 2012 Jan;48(1):37-44. [doi: 10.1016/j.jesp.2011.07.001]
17. Rovira A, Swapp D, Spanlang B, Slater M. The use of virtual reality in the study of people's responses to violent incidents. Front Behav Neurosci 2009;3:59. [doi: 10.3389/neuro.08.059.2009] [Medline: 20076762]
18. Fisher S, Abdullah A, Charvin I, Da Fonseca D, Bat-Pitault F. Comparison of body image evaluation by virtual reality and paper-based figure rating scales in adolescents with anorexia nervosa: retrospective study. Eat Weight Disord 2020 Jun;25(3):735-743. [doi: 10.1007/s40519-019-00680-1] [Medline: 30977098]
19. Pan X, Hamilton ADC. Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. Br J Psychol 2018 Aug;109(3):395-417. [doi: 10.1111/bjop.12290] [Medline: 29504117]
20. Ponton JL, Yun H, Andujar C, Pelechano N. Combining motion matching and orientation prediction to animate avatars for consumer‐grade VR devices. Comput Graph Forum 2022 Dec;41(8):107-118. [doi: 10.1111/cgf.14628]
21. Tsoli A, Mahmood N, Black MJ. Breathing life into shape: capturing, modeling and animating 3D human breathing. ACM Trans Graph 2014 Jul 27;33(4):1-11. [doi: 10.1145/2601097.2601225]
22. Zibrek K, Kokkinara E, Mcdonnell R. The effect of realistic appearance of virtual characters in immersive environments - does the character's personality play a role? IEEE Trans Vis Comput Graph 2018 Apr;24(4):1681-1690. [doi: 10.1109/TVCG.2018.2794638] [Medline: 29543183]
23. Llanes-Jurado J, Gómez-Zaragozá L, Minissi ME, Alcañiz M, Marín-Morales J. Developing conversational virtual humans for social emotion elicitation based on large language models. Expert Syst Appl 2024 Jul;246:123261. [doi: 10.1016/j.eswa.2024.123261]
24. Rovira A, Slater M. Encouraging bystander helping behaviour in a violent incident: a virtual reality study using reinforcement learning. Sci Rep 2022 Mar 9;12(1):3843. [doi: 10.1038/s41598-022-07872-3] [Medline: 35264652]
25. Wienrich C, Latoschik ME. Extended artificial intelligence: new prospects of human-AI interaction research. Front Virtual Real 2021 Sep 6;2. [doi: 10.3389/frvir.2021.686783]
26. Gonzalez-Franco M, Egan Z, Peachey M, et al. MoveBox: democratizing MoCap for the Microsoft rocketbox avatar library. Presented at: 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR); Dec 14-18, 2020; Utrecht, Netherlands. [doi: 10.1109/AIVR50618.2020.00026]
27. Gonzalez-Franco M, Ofek E, Pan Y, et al. The rocketbox library and the utility of freely available rigged avatars. Front Virtual Real 2020 Nov 3;1. [doi: 10.3389/frvir.2020.561558]
28. Volonte M, Ofek E, Jakubzak K, Bruner S, Gonzalez-Franco M. HeadBox: a facial blendshape animation toolkit for the microsoft rocketbox library. Presented at: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); Mar 12-16, 2022; Christchurch, New Zealand. [doi: 10.1109/VRW55335.2022.00015]
29. Chattopadhyay D, Ma T, Sharifi H, Martyn-Nemeth P. Computer-controlled virtual humans in patient-facing systems: systematic review and meta-analysis. J Med Internet Res 2020 Jul 30;22(7):e18839. [doi: 10.2196/18839] [Medline: 32729837]
30. Vinayagamoorthy V, Gillies M, Steed A, et al. Building expression into virtual characters. Presented at: Eurographics 2006; Sep 4-8, 2006; Vienna, Austria URL: https://eprints.gold.ac.uk/id/eprint/398/1/expressivevirtualcharacters.pdf [accessed 2025-09-28]
31. Hameed A, Möller S, Perkis A. A holistic quality taxonomy for virtual reality experiences. Front Virtual Real 2024 Aug 30;5. [doi: 10.3389/frvir.2024.1434016]
32. Arsenyan J, Mirowska A, Piepenbrink A. Close encounters with the virtual kind: defining a human-virtual agent coexistence framework. Technol Forecast Soc Change 2023 Aug;193:122644. [doi: 10.1016/j.techfore.2023.122644]
33. Kyrlitsias C, Michael-Grigoriou D. Social interaction with agents and avatars in immersive virtual environments: a survey. Front Virtual Real 2022 Jan 11;2. [doi: 10.3389/frvir.2021.786665]

34. Pimentel D, Vinkers C. Copresence with virtual humans in mixed reality: the impact of contextual responsiveness on social perceptions. Front Robot AI 2021;8:634520. [doi: 10.3389/frobt.2021.634520] [Medline: 33912595]
35. Blascovich J, Bailenson J. Infinite Reality: Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution: William Morrow & Co; 2011.
36. Volonte M, Hsu YC, Liu KY, Mazer JP, Wong SK, Babu SV. Effects of interacting with a crowd of emotional virtual humans on users' affective and non-verbal behaviors. Presented at: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); Mar 22-26, 2020; Atlanta, GA, USA. [doi: 10.1109/VR46266.2020.00049]
37. Wei S, Freeman D, Rovira A. A randomised controlled test of emotional attributes of a virtual coach within a virtual reality (VR) mental health treatment. Sci Rep 2023 Jul 17;13(1):11517. [doi: 10.1038/s41598-023-38499-7] [Medline: 37460586]
38. Jorjafki EM, Sagarin BJ, Butail S. Drawing power of virtual crowds. J R Soc Interface 2018 Aug;15(145):20180335. [doi: 10.1098/rsif.2018.0335] [Medline: 30111664]
39. Rizzo AS, Difede J, Rothbaum BO, et al. Development and early evaluation of the Virtual Iraq/Afghanistan exposure therapy system for combat-related PTSD. Ann N Y Acad Sci 2010 Oct;1208(1):114-125. [doi: 10.1111/j.1749-6632.2010.05755.x] [Medline: 20955333]
40. Ferrer-Garcia M, Porras-Garcia B, Gonzalez-Ibanez C, et al. Does owning a "fatter" virtual body increase body anxiety in college students? Annu Rev Cyberther Telemed 2017;15:147-153 [FREE Full text]
41. Nowak KL, Biocca F. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. Presence: Teleoperators & Virtual Environments 2003 Oct;12(5):481-494. [doi: 10.1162/105474603322761289]
42. Cassell J. Embodied conversational agents: representation and intelligence in user interfaces. AI Mag 2001 Dec 15;22(4):67-67. [doi: 10.1609/aimag.v22i4.1593]
43. Salber D, Coutaz J. Applying the Wizard of Oz technique to the study of multimodal systems. In: Bass LJ, Gornostaev J, Unger C, editors. Human-Computer Interaction. EWHCI 1993. Lecture Notes in Computer Science: Springer; 1993:219-230. [doi: 10.1007/3-540-57433-6_51]
44. Joundi J, Bombeke K, Kets N, et al. ExperienceDNA: a framework to conduct and analyse user tests in VR using the Wizard-of-Oz methodology. In: Soares MM, Rosenzweig E, Marcus A, editors. Design, User Experience, and Usability: Design for Contemporary Technological Environments HCII 2021 Lecture Notes in Computer Science:171-186. [doi: 10.1007/978-3-030-78227-6_13]
45. Kothgassner OD, Goreis A, Glenk LM, et al. Virtual and real-life ostracism and its impact on a subsequent acute stressor. Physiol Behav 2021 Jan 1;228(113205):113205. [doi: 10.1016/j.physbeh.2020.113205] [Medline: 33058901]
46. Slater M, Wilbur S. A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. Presence: Teleoperators & Virtual Environments 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
47. Spyridonis F, Daylamani-Zad D, Nightingale J. PublicVR: a virtual reality exposure therapy intervention for adults with speech anxiety. Virtual Real 2024;28(2). [doi: 10.1007/s10055-024-00998-x]
48. Brinkman WP, Veling W, Dorrestijn E, Sandino G, Vakili V, van der Gaag M. Virtual reality to study responses to social environmental stressors in individuals with and without psychosis. Stud Health Technol Inform 2011;167:86-91. [doi: 10.3233/978-1-6-750-766-6-86] [Medline: 21685647]
49. Tarnanas I, Tsoukalas I, Stogiannidou A. Virtual reality as a psychosocial coping environment. In: Wiederhold BK, Riva G, editors. Annual Review of CyberTherapy and Telemedicine: Interactive Media Institute; 2003, Vol. 1:77-81 URL: https://www.arctt.info/volume-1-summer-2003 [accessed 2025-09-27]
50. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160. [doi: 10.1136/bmj.n160] [Medline: 33781993]
51. Lee J, Lim Y, Graham SJ, et al. Nicotine craving and cue exposure therapy by using virtual environments. Cyberpsychol Behav 2004 Dec;7(6):705-713. [doi: 10.1089/cpb.2004.7.705] [Medline: 15687806]
52. Park IH, Kim JJ, Jang HJ, et al. Characteristics of social anxiety from virtual interpersonal interactions in patients with schizophrenia. Psychiatry: Interpersonal and Biological Processes 2009 Mar;72(1):79-93. [doi: 10.1521/psyc.2009.72.1.79]
53. Wieser MJ, Pauli P, Grosseibl M, Molzow I, Mühlberger A. Virtual social interactions in social anxiety--the impact of sex, gaze, and interpersonal distance. Cyberpsychol Behav Soc Netw 2010 Oct;13(5):547-554. [doi: 10.1089/cyber.2009.0432] [Medline: 20950179]
54. Kwon JH, Powell J, Chalmers A. How level of realism influences anxiety in virtual reality environments for a job interview. Int J Hum Comput Stud 2013 Oct;71(10):978-987. [doi: 10.1016/j.ijhcs.2013.07.003]
55. Powers MB, Briceno NF, Gresham R, Jouriles EN, Emmelkamp PMG, Smits JAJ. Do conversations with virtual avatars increase feelings of social anxiety? J Anxiety Disord 2013 May;27(4):398-403. [doi: 10.1016/j.janxdis.2013.03.003] [Medline: 23746490]
56. Han K, Shin J, Yoon SY, Jang DP, Kim JJ. Deficient gaze pattern during virtual multiparty conversation in patients with schizophrenia. Comput Biol Med 2014 Jun;49:60-66. [doi: 10.1016/j.compbiomed.2014.03.012] [Medline: 24747729]
57. Hartanto D, Kampmann IL, Morina N, Emmelkamp PGM, Neerincx MA, Brinkman WP. Controlling social stress in virtual reality environments. PLOS ONE 2014;9(3):e92804. [doi: 10.1371/journal.pone.0092804] [Medline: 24671006]

58. Pan X, Gillies M, Slater M. Virtual character personality influences participant attitudes and behavior - an interview with a virtual human character about her social anxiety. Front Robot AI 2015;2. [doi: 10.3389/frobt.2015.00001]
59. Ryan V, Griffin R. An investigation into anxiety in virtual reality following a self-compassion induction. In: Wiederhold BK, Riva G, Wiederhold MD, Kirwan G, editors. Annual Review of CyberTherapy and Telemedicine 2016, Vol. 14:109-114.
60. Robitaille N, Jackson PL, Hébert LJ, et al. A virtual reality avatar interaction (VRai) platform to assess residual executive dysfunction in active military personnel with previous mild traumatic brain injury: proof of concept. Disabil Rehabil Assist Technol 2017 Oct;12(7):758-764. [doi: 10.1080/17483107.2016.1229048] [Medline: 27677827]
61. Percie du Sert O, Potvin S, Lipp O, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: a pilot clinical trial. Schizophr Res 2018 Jul;197:176-181. [doi: 10.1016/j.schres.2018.02.031] [Medline: 29486956]
62. Freeman D, Haselton P, Freeman J, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. Lancet Psychiatry 2018 Aug;5(8):625-632. [doi: 10.1016/S2215-0366(18)30226-8] [Medline: 30007519]
63. Shin YB, Kim JJ, Kim MK, et al. Development of an effective virtual environment in eliciting craving in adolescents and young adults with internet gaming disorder. PLOS ONE 2018;13(4):e0195677. [doi: 10.1371/journal.pone.0195677] [Medline: 29672530]
64. Kothgassner OD, Goreis A, Kafka JX, et al. Virtual social support buffers stress response: an experimental comparison of real-life and virtual support prior to a social stressor. J Behav Ther Exp Psychiatry 2019 Jun;63:57-65. [doi: 10.1016/j.jbtep.2018.11.003] [Medline: 30454876]
65. Quintana P, Nolet K, Baus O, Bouchard S. The effect of exposure to fear-related body odorants on anxiety and interpersonal trust toward a virtual character. Chem Senses 2019 Oct 26;44(9):683-692. [doi: 10.1093/chemse/bjz063] [Medline: 31504295]
66. Reichenberger J, Pfaller M, Forster D, Gerczuk J, Shiban Y, Mühlberger A. Men scare me more: gender differences in social fear conditioning in virtual reality. Front Psychol 2019;10:1617. [doi: 10.3389/fpsyg.2019.01617] [Medline: 31417443]
67. Seo SH, Kim E, Mundy P, Heo J, Kim KK. Joint attention virtual classroom: a preliminary study. Psychiatry Investig 2019 Apr;16(4):292-299. [doi: 10.30773/pi.2019.02.08]
68. Miloff A, Carlbring P, Hamilton W, Andersson G, Reuterskiöld L, Lindner P. Measuring alliance toward embodied virtual therapists in the era of automated treatments with the virtual therapist alliance scale (VTAS): development and psychometric evaluation. J Med Internet Res 2020 Mar 24;22(3):e16660. [doi: 10.2196/16660] [Medline: 32207690]
69. Nijman SA, Veling W, Greaves-Lord K, et al. Dynamic interactive social cognition training in virtual reality (DiSCoVR) for people with a psychotic disorder: single-group feasibility and acceptability study. JMIR Ment Health 2020 Aug 7;7(8):e17808. [doi: 10.2196/17808] [Medline: 32763880]
70. Kim J, Jung YH, Shin YB, et al. Development and validation of a virtual reality-based training program for promoting subjective well-being. Psychiatry Investig 2020 Dec;17(12):1207-1215. [doi: 10.30773/pi.2020.0311]
71. Lee HS, Hong SJJ, Baxter T, et al. Altered peripersonal space and the bodily self in schizophrenia: a virtual reality study. Schizophr Bull 2021 Jul 8;47(4):927-937. [doi: 10.1093/schbul/sbab024] [Medline: 33844019]
72. Brander M, Egger ST, Hürlimann N, et al. Virtual reality human-human interface to deliver psychotherapy to people experiencing auditory verbal hallucinations: development and usability study. JMIR Serious Games 2021 Jun 1;9(2):e26820. [doi: 10.2196/26820] [Medline: 33769295]
73. Guldager JD, Kjær SL, Grittner U, Stock C. Efficacy of the virtual reality intervention VR FestLab on alcohol refusal self-efficacy: a cluster-randomized controlled trial. Int J Environ Res Public Health 2022 Mar 10;19(6):3293. [doi: 10.3390/ijerph19063293] [Medline: 35328977]
74. Bektas S, Natali L, Rowlands K, et al. Exploring correlations of food-specific disgust with eating disorder psychopathology and food interaction: a preliminary study using virtual reality. Nutrients 2023 Oct 19;15(20):4443. [doi: 10.3390/nu15204443] [Medline: 37892518]
75. Giguère S, Potvin S, Beaudoin M, et al. Avatar intervention for cannabis use disorder in individuals with severe mental disorders: a pilot study. J Pers Med 2023 Apr 29;13(5):766. [doi: 10.3390/jpm13050766] [Medline: 37240936]
76. Artiran S, Bedmutha PS, Cosman P. Analysis of gaze, head orientation, and joint attention in autism with triadic VR interviews. IEEE Trans Neural Syst Rehabil Eng 2024;32:759-769. [doi: 10.1109/TNSRE.2024.3363728] [Medline: 38329870]
77. Natali L, Meregalli V, Rowlands K, et al. Virtual food exposure with positive mood induction or social support to reduce food anxiety in anorexia nervosa: a feasibility study. Int J Eat Disord 2024 Mar;57(3):703-715. [doi: 10.1002/eat.24155] [Medline: 38366755]
78. Hidding M, Veling W, Pijnenborg GHM, van der Stouwe ECD. A single-session VR intervention addressing self-compassion and self-criticism with and without perspective change: results of a randomized controlled experiment. Behav Res Ther 2024 Feb;173:104466. [doi: 10.1016/j.brat.2023.104466] [Medline: 38141543]
79. Yamashita Y, Yamamoto T. Effect of virtual reality self-counseling with the intimate other avatar. Sci Rep 2024 Jul 4;14(1):15417. [doi: 10.1038/s41598-024-65661-6] [Medline: 38965334]
80. Banakou D, Johnston T, Beacco A, Senel G, Slater M. Desensitizing anxiety through imperceptible change: feasibility study on a paradigm for single-session exposure therapy for fear of public speaking. JMIR Form Res 2024 Jul 22;8(1):e52212. [doi: 10.2196/52212] [Medline: 39037760]

81. Freeman D, Slater M, Bebbington PE, et al. Can virtual reality be used to investigate persecutory ideation? J Nerv Ment Dis 2003 Aug;191(8):509-514. [doi: 10.1097/01.nmd.0000082212.83842.fe] [Medline: 12972853]
82. Freeman D, Garety PA, Bebbington P, et al. The psychology of persecutory ideation II: a virtual reality experimental study. J Nerv Ment Dis 2005 May;193(5):309-315. [doi: 10.1097/01.nmd.0000161686.53245.70] [Medline: 15870614]
83. Cho S, Ku J, Park J, et al. Development and verification of an alcohol craving-induction tool using virtual reality: craving characteristics in social pressure situation. Cyberpsychol Behav 2008 Jun;11(3):302-309. [doi: 10.1089/cpb.2007.0149] [Medline: 18537500]
84. Shiban Y, Reichenberger J, Neumann ID, Mühlberger A. Social conditioning and extinction paradigm: a translational study in virtual reality. Front Psychol 2015;6(400):400. [doi: 10.3389/fpsyg.2015.00400] [Medline: 25904889]
85. Mountford VA, Tchanturia K, Valmaggia L. "What are you thinking when you look at me?" A pilot study of the use of virtual reality in body image. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):93-99. [doi: 10.1089/cyber.2015.0169] [Medline: 26378613]
86. Atherton S, Antley A, Evans N, et al. Self-confidence and paranoia: an experimental study using an immersive virtual reality social situation. Behav Cogn Psychother 2016 Jan;44(1):56-64. [doi: 10.1017/S1352465814000496] [Medline: 25384608]
87. Takac M, Collett J, Blom KJ, Conduit R, Rehm I, De Foe A. Public speaking anxiety decreases within repeated virtual reality training sessions. PLoS ONE 2019;14(5):e0216288. [doi: 10.1371/journal.pone.0216288] [Medline: 31150399]
88. Wei S, Freeman D, Harris V, Rovira A. A randomised controlled test in virtual reality of the effects on paranoid thoughts of virtual humans' facial animation and expression. Sci Rep 2024 Jul 24;14(1):17102. [doi: 10.1038/s41598-024-67534-4]
89. Gayer-Anderson C, Knowles G, Beards S, et al. Immersive virtual reality as a novel approach to investigate the association between adverse events and adolescent paranoid ideation. Soc Psychiatry Psychiatr Epidemiol 2025 Feb;60(2):305-318. [doi: 10.1007/s00127-024-02701-6] [Medline: 38942902]
90. Rovira A, Lambe S, Beckwith H, et al. A randomized controlled experiment testing the use of virtual reality to trigger cigarette craving in people who smoke. Sci Rep 2024 Aug 21;14(1):19445. [doi: 10.1038/s41598-024-70113-2] [Medline: 39169108]
91. Perpiñá C, Botella C, Baños R, Marco H, Alcañiz M, Quero S. Body image and virtual reality in eating disorders: is exposure to virtual reality more effective than the classical body image treatment? Cyberpsychol Behav 1999;2(2):149-155. [doi: 10.1089/cpb.1999.2.149] [Medline: 19178251]
92. Aymerich-Franch L, Kizilcec RF, Bailenson JN. The relationship between virtual self similarity and social anxiety. Front Hum Neurosci 2014;8(944):944. [doi: 10.3389/fnhum.2014.00944] [Medline: 25477810]
93. Ferrer-Garcia M, Porras-Garcia B, Moreno M, Bertomeu P, Maldonado JG. Embodiment in different size virtual bodies produces changes in women's body image distortion and dissatisfaction. In: Wiederhold BK, Riva G, Bouchard S, editors. Annual Review of CyberTherapy and Telemedicine 2018, Vol. 16:111-117 URL: https://www.arctt.info/volume-16-summer-2018 [accessed 2025-09-27]
94. Porras-Garcia B, Ghita A, Moreno M, et al. Gender differences in attentional bias after owning a virtual avatar with increased weight. In: Wiederhold BK, Riva G, Bouchard S, editors. Annual Review of CyberTherapy and Telemedicine 2018, Vol. 16:73-79.
95. Porras-Garcia B, Ferrer-Garcia M, Ghita A, et al. The influence of gender and body dissatisfaction on body-related attentional bias: an eye-tracking and virtual reality study. Int J Eat Disord 2019 Oct;52(10):1181-1190. [doi: 10.1002/eat.23136] [Medline: 31301151]
96. Porras Garcia B, Ferrer Garcia M, Olszewska A, et al. Is this my own body? Changing the perceptual and affective body image experience among college students using a new virtual reality embodiment-based technique. J Clin Med 2019 Jun 27;8(7):925. [doi: 10.3390/jcm8070925] [Medline: 31252596]
97. Mertens G, Wagensveld P, Engelhard IM. Cue conditioning using a virtual spider discriminates between high and low spider fearful individuals. Comput Human Behav 2019 Feb;91:192-200. [doi: 10.1016/j.chb.2018.10.006]
98. Provenzano L, Porciello G, Ciccarone S, et al. Characterizing body image distortion and bodily self-plasticity in anorexia nervosa via visuo-tactile stimulation in virtual reality. J Clin Med 2020 Jan;9(1):98. [doi: 10.3390/jcm9010098]
99. Scarpina F, Serino S, Keizer A, et al. The effect of a virtual-reality full-body illusion on body representation in obesity. J Clin Med 2019 Aug 28;8(9):1330. [doi: 10.3390/jcm8091330] [Medline: 31466376]
100. Spanlang B, Nierula B, Haffar M, Debruille JB. Mimicking schizophrenia: reducing P300b by minimally fragmenting healthy participants' selves using immersive virtual reality embodiment. Front Hum Neurosci 2018;12(504):504. [doi: 10.3389/fnhum.2018.00504] [Medline: 30662398]
101. Porras‐Garcia B, Ferrer‐Garcia M, Yilmaz L, et al. Body‐related attentional bias as mediator of the relationship between body mass index and body dissatisfaction. Euro Eating Disorders Rev 2020 Jul;28(4):454-464. [doi: 10.1002/erv.2730]
102. Gorisse G, Senel G, Banakou D, et al. Self-observation of a virtual body-double engaged in social interaction reduces persecutory thoughts. Sci Rep 2021 Dec 14;11(1):23923. [doi: 10.1038/s41598-021-03373-x] [Medline: 34907279]
103. Wolf E, Merdan N, Dolinger N, et al. The embodiment of photorealistic avatars influences female body weight perception in virtual reality. Presented at: 2021 IEEE Virtual Reality and 3D User Interfaces (VR); Mar 27 to Apr 1, 2021; Lisboa, Portugal. [doi: 10.1109/VR50410.2021.00027]

104. van Gelder JL, Cornet LJM, Zwalua NP, Mertens ECA, van der Schalk J. Interaction with the future self in virtual reality reduces self-defeating behavior in a sample of convicted offenders. Sci Rep 2022 Feb 10;12(1):2254. [doi: 10.1038/s41598-022-06305-5] [Medline: 35145177]
105. Yamamoto K, Nakao T. Top-down influence leads to a reduced sense of body ownership in individuals with depersonalization tendencies: a focus on full body illusion. Front Psychol 2022;13:790960. [doi: 10.3389/fpsyg.2022.790960] [Medline: 35719590]
106. Burin D, Cavanna G, Rabellino D, Kotozaki Y, Kawashima R. Neuroendocrine response and state anxiety due to psychosocial stress decrease after a training with subject's own (but not another) virtual body: an RCT study. Int J Environ Res Public Health 2022 May 23;19(10):6340. [doi: 10.3390/ijerph19106340] [Medline: 35627877]
107. Vahle N, Tomasik MJ. The embodiment of an older avatar in a virtual reality setting impacts the social motivation of young adults. Exp Aging Res 2022;48(2):164-176. [doi: 10.1080/0361073X.2021.1943793] [Medline: 34182895]
108. Mendoza-Medialdea MT, Meschberger-Annweiler FA, Ascione M, et al. Body dissatisfaction and body-related attentional bias: is there a causal relationship? J Clin Med 2023 Aug 30;12(17):5659. [doi: 10.3390/jcm12175659] [Medline: 37685726]
109. Ascione M, Carulla-Roig M, Miquel-Nabau H, et al. Attentional bias modification training based on virtual reality and eye tracking in anorexia nervosa patients. J Clin Med 2023 Sep 12;12(18):5932. [doi: 10.3390/jcm12185932] [Medline: 37762873]
110. Schroeder PA, Gehrer NA, Reents M, Reimer N, Vagedes J, Svaldi J. Body dissatisfaction directs avatar perception: embodiment and selective visual attention in body mass-modified self-avatars. Cyberpsychol Behav Soc Netw 2023 Nov;26(11):850-860. [doi: 10.1089/cyber.2022.0385] [Medline: 37948539]
111. Wei S, Freeman D, Rovira A. Visual attention and virtual human facial animations in virtual reality (VR): an eye-tracking study. Presented at: 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); Mar 16-21, 2024; Orlando, FL, USA. [doi: 10.1109/VRW62533.2024.00242]
112. Magnenat-Thalmann N, Thalmann D. Virtual humans: thirty years of research, what next? Visual Comput 2005 Dec;21(12):997-1015. [doi: 10.1007/s00371-005-0363-6]
113. Zell E, Zibrek K, McDonnell R. Perception of virtual characters. Presented at: SIGGRAPH '19: Special Interest Group on Computer Graphics and Interactive Techniques Conference; Jul 28, 2019; Los Angeles, CA, USA. [doi: 10.1145/3305366.3328101]
114. Bombari D, Schmid Mast M, Canadas E, Bachmann M. Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. Front Psychol 2015;6:869. [doi: 10.3389/fpsyg.2015.00869] [Medline: 26157414]
115. Gillies M. Creating virtual characters. Presented at: MOCO '18: 5th International Conference on Movement and Computing; Jun 28-30, 2018; Genoa, Italy. [doi: 10.1145/3212721.3212835]
116. Felnhofer A, Knaust T, Weiss L, Goinska K, Mayer A, Kothgassner OD. A virtual character's agency affects social responses in immersive virtual reality: a systematic review and meta-analysis. Int J Hum Comput Int 2024;40(16):4167-4182. [doi: 10.1080/10447318.2023.2209979]
117. Stein JP, Ohler P. Venturing into the uncanny valley of mind: the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. Cognition 2017 Mar;160:43-50. [doi: 10.1016/j.cognition.2016.12.010] [Medline: 28043026]
118. Choudhary Z, Norouzi N, Erickson A, Schubert R, Bruder G, Welch GF. Exploring the social influence of virtual humans unintentionally conveying conflicting emotions. Presented at: 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR); Mar 25-29, 2023; Shanghai, China. [doi: 10.1109/VR55154.2023.00072]
119. Dobre GC, Gillies M, Pan X. Immersive machine learning for social attitude detection in virtual reality narrative games. Virtual Real 2022 Dec;26(4):1519-1538. [doi: 10.1007/s10055-022-00644-4]
120. Gupta K, Zhang Y, Gunasekaran TS, Krishna N, Pai YS, Billinghurst M. CAEVR: biosignals-driven context-aware empathy in virtual reality. IEEE Trans Visual Comput Graphics 2024;30(5):2671-2681. [doi: 10.1109/TVCG.2024.3372130]
121. Venning A, Herd MC, Oswald TK, et al. Exploring the acceptability of a digital mental health platform incorporating a virtual coach: the good, the bad, and the opportunities. Health Informatics J 2021;27(1):1460458221994873. [doi: 10.1177/1460458221994873] [Medline: 33601947]
122. Knight I, West J, Matthews E, et al. Participatory design to create a VR therapy for psychosis. Design for Health 2021 May;5(1):98-119. [doi: 10.1080/24735132.2021.1885889]
123. Lambe S, Knight I, Kabir T, et al. Developing an automated VR cognitive treatment for psychosis: gameChange VR therapy. J Behav Cogn Ther 2020 Apr;30(1):33-40. [doi: 10.1016/j.jbct.2019.12.001]
124. Dickinson P, Gerling K, Hicks K, Murray J, Shearer J, Greenwood J. Virtual reality crowd simulation: effects of agent density on user experience and behaviour. Virtual Real 2019 Mar 5;23(1):19-32. [doi: 10.1007/s10055-018-0365-0]
125. Ulicny B, Thalmann D. Crowd simulation for interactive virtual environments and VR training systems. In: Magnenat-Thalmann N, Thalmann D, editors. Computer Animation and Simulation 2001: Springer; 2001:163-170. [doi: 10.1007/978-3-7091-6240-8_15]
126. Dong Y, Peng C. Real-time large crowd rendering with efficient character and instance management on GPU. Int J Comput Games Technol 2019 Mar 26;2019:1-15. [doi: 10.1155/2019/1792304]

127. Magnenat-Thalmann N. Challenges for the research in virtual humans. 2000 URL: https://www.academia.edu/76018642/Challenges_for_the_research_in_virtual_humans [accessed 2025-09-27]
128. Trivedi H, Mousas C. Human‐virtual crowd interaction: towards understanding the effects of crowd avoidance proximity in an immersive virtual environment. Computer Animation and Virtual Worlds 2023 May;34(3-4). [doi: 10.1002/cav.2169]
129. Jain D, Jain U, Rao K, et al. A real-time crowd tracking and control system using deep learning. In: Hassanien AE, Anand S, Jaiswal A, Kumar P, editors. Innovative Computing and Communications. ICICC 2024. Lecture Notes in Networks and Systems: Springer; 2024:587-601. [doi: 10.1007/978-981-97-3817-5_42]
130. Cerritelli F, Chiera M, Abbro M, et al. The challenges and perspectives of the integration between virtual and augmented reality and manual therapies. Front Neurol 2021;12(700211):700211. [doi: 10.3389/fneur.2021.700211] [Medline: 34276550]
131. Spanlang B, Normand JM, Borland D, et al. How to build an embodiment lab: achieving body representation illusions in virtual reality. Front Robot AI 2014;1:9. [doi: 10.3389/frobt.2014.00009]
132. Flavián C, Ibáñez-Sánchez S, Orús C. The influence of scent on virtual reality experiences: the role of aroma-content congruence. J Bus Res 2021 Feb;123:289-301. [doi: 10.1016/j.jbusres.2020.09.036]
133. Leonardis D, Frisoli A, Barsotti M, Carrozzino M, Bergamasco M. Multisensory feedback can enhance embodiment within an enriched virtual walking scenario. Presence: Teleoperators and Virtual Environments 2014 Oct 1;23(3):253-266. [doi: 10.1162/PRES_a_00190]
134. Banakou D, Beacco A, Neyret S, Blasco-Oliver M, Seinfeld S, Slater M. Virtual body ownership and its consequences for implicit racial bias are dependent on social context. R Soc Open Sci 2020 Dec;7(12):201848. [doi: 10.1098/rsos.201848] [Medline: 33489296]
135. Kao D, Ratan R, Mousas C, Magana AJ. The effects of a self-similar avatar voice in educational games. Proc ACM Hum-Comput Interact 2021 Oct 5;5(CHI PLAY):1-28. [doi: 10.1145/3474665]
136. Evidence standards framework for digital health technologies. NICE. 2018. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2025-09-27]
137. Eysenbach G, Group CE. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
138. Vlake JH, Drop DLQ, Van Bommel J, et al. Reporting guidelines for the early-phase clinical evaluation of applications using extended reality: RATE-XR qualitative study guideline. J Med Internet Res 2024 Nov 29;26(1):e56790. [doi: 10.2196/56790] [Medline: 39612482]
139. Wei S, El Ali A, Cesar P, Freeman D, Rovira A. Physiological responses to affective virtual coach design in a VR fear of heights consultation. Presented at: CHI EA '25: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems; Apr 26 to May 1, 2025; Yokohama, Japan. [doi: 10.1145/3706599.3719715]
140. Kumarapeli D, Jung S, Lindeman RW. Privacy threats of behaviour identity detection in VR. Front Virtual Real 2024 Jan 29;5. [doi: 10.3389/frvir.2024.1197547]
141. Döllinger N, Wolf E, Botsch M, Latoschik ME, Wienrich C. Are embodied avatars harmful to our self-experience? The impact of virtual embodiment on body awareness. Presented at: CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 23-28, 2023; Hamburg, Germany. [doi: 10.1145/3544548.3580918]
142. de Villiers Bosman I, Bujić M, Cosio LD, 'Oz' Buruk O, Jørgensen K, Hamari J. Embodiment in virtual reality: an experiment on how visual and aural first- and third-person modes affect embodiment and mindfulness. Virtual Real 2025;29(2). [doi: 10.1007/s10055-025-01129-w]
143. Saberi M, Bernardet U. An architecture for personality-based, nonverbal behavior in affective virtual humanoid character. Procedia Comput Sci 2014 Jan 1;41:204-211. [doi: 10.1016/j.procs.2014.11.104]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
VA: virtual agent
VH: virtual human
VR: virtual reality

Edited by A Coristine; submitted 28.03.25; peer-reviewed by A Azmi, JF Costa; revised version received 21.09.25; accepted 22.09.25; published 08.10.25.

Please cite as:
Wei S, Freeman D, Rovira A
Virtual Humans in Virtual Reality Mental Health Research: Systematic Review
JMIR XR Spatial Comput 2025;2:e75087
URL: https://xr.jmir.org/2025/1/e75087
doi:10.2196/75087

Virtual Reality Reminiscence Therapy in Dementia Care: Scoping Review of Research

Maria Matsangidou, MA, MA, PhD

CYENS Centre of Excellence, Lellou Demetriades, Plateia Dimarchou 1, Nicosia, Cyprus

Corresponding Author:
Maria Matsangidou, MA, MA, PhD
CYENS Centre of Excellence, Lellou Demetriades, Plateia Dimarchou 1, Nicosia, Cyprus

Abstract

Background: Dementia is a progressive neurological disorder affecting cognitive and social functioning, posing challenges for patients and caregivers. Traditional medications often have adverse effects, emphasizing the need for nonpharmacological options such as reminiscence therapy (RT). Virtual reality (VR) has emerged as a promising tool in dementia care, providing immersive experiences that stimulate memory, enhance emotional well-being, and reduce the behavioral and psychological symptoms of dementia.

Objective: This scoping review assesses the feasibility and implementation challenges of delivering RT via VR in dementia care. Specifically, it examines the types of VR systems used, their therapeutic benefits, and the barriers to their adoption.

Methods: We screened 5 electronic libraries: Google Scholar, ACM Digital Library, IEEE Xplore, MEDLINE, and PubMed. Studies published between 2000 and 2025 were included if they examined the use of VR for RT in people with dementia. Data were charted based on PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines and analyzed thematically for feasibility, VR system type, therapeutic effects, and implementation considerations.

Results: A total of 15 studies met the inclusion criteria. The findings indicate that VR is feasible and well-accepted among people with dementia, fostering high engagement with minimal adverse effects. Fully immersive VR systems, which use head-mounted displays, are the most frequently used, while semi-immersive alternatives with large screens provide a more cost-effective option. RT via VR has been shown to improve reminiscence, enhance mood, and encourage social interaction. However, its impact on cognitive function remains inconclusive. Significant barriers to implementation include high costs, limited availability of VR infrastructure in care, and the need for specialized caregiver training.

Conclusions: RT via VR presents a promising advancement in dementia care. Future research should focus on developing cost-effective, scalable VR solutions, designing personalized VR experiences tailored to individual needs, and creating structured training programs for caregivers. Longitudinal studies are necessary to determine the long-term therapeutic effects of VR compared to traditional RT.

(JMIR XR Spatial Comput 2025;2:e73539) doi:10.2196/73539

KEYWORDS

dementia; virtual reality; reminiscence therapy; nonpharmacological interventions; cognitive stimulation; emotional stimulation

Introduction

Background

Dementia is a progressive neurological disorder and an umbrella term for conditions that lead to the deterioration of cognitive and social functioning. It impacts various mental faculties, including memory, problem-solving, orientation, comprehension, calculation, learning capacity, language, and judgment [1]. People with dementia often experience behavioral and psychological symptoms of dementia (BPSD), including mood disturbances, aggression, irritability, apathy, and emotional dysregulation, which further reduce their quality of life and place significant burdens on caregivers [2,3].

Worldwide, dementia affects a rapidly growing population. Only in 2021, the World Health Organization estimated that over 55 million people were living with dementia, a number expected to rise to 82 million by 2030 and 139 million by 2050 [1]. It is currently the seventh leading cause of death worldwide and a major cause of disability and dependency among older adults [1]. This rising prevalence underscores the urgent need for interventions, as dementia imposes significant physical, psychological, social, and economic burdens [2].

While no cure exists, treatments aim to alleviate symptoms and improve the quality of life for people with dementia. Pharmacological treatments, such as neuroleptic or sedating medications, have historically been overused despite their

association with adverse effects, including accelerated cognitive decline, cardiovascular issues, infections, and emotional distress [4,5]. Consequently, nonpharmacological interventions have gained prominence as preferred approaches for addressing dementia symptoms. Among these, reminiscence therapy (RT) has emerged as a particularly effective psychosocial intervention. RT involves structured recall and discussion of past experiences using prompts such as photographs, music, personal objects, and storytelling [6]. Evidence suggests that RT can improve mood, reduce agitation, and enhance social interaction in people with dementia [6,7].

The application of RT varies significantly across studies, with interventions delivered in both personalized and nonpersonalized formats. Some studies combined both approaches [8], while others focused exclusively on either personalized RT [9,10] or nonpersonalized RT [6,11]. RT methods, such as group storytelling, memory books, multisensory activities, and music therapy, have been widely used [9,12]. Meanwhile, digital RT, which integrates digital apps, video games, and web-based platforms, has introduced novel ways of engaging people with dementia [8,10,11]. Such digital interventions have shown promise in increasing engagement and reducing social isolation [12]. However, challenges such as cost, feasibility, and inconsistent therapeutic outcomes remain [6].

Emerging Role of Virtual Reality in RT

Recent technological advancements, particularly virtual reality (VR), have expanded the possibilities for nonpharmacological dementia care [13,14]. VR creates immersive, computer-generated environments that simulate real-world experiences, engaging users' senses through visual, auditory, tactile, and even olfactory stimuli [15]. These environments enable people with dementia to transcend physical limitations and explore familiar or calming settings, such as natural landscapes or culturally significant landmarks [14,16,17]. By simulating personalized environments, VR interventions have demonstrated the potential to reduce anxiety, depression, aggression, and social withdrawal while promoting cognitive engagement and emotional well-being [13,14,18-20].

Studies show that personalized VR experiences, tailored to an individual's preferences and abilities, have been linked to improved cognitive functions, such as memory recall and spatial navigation, as well as reductions in agitation and emotional dysregulation [21]. Moreover, VR-based interventions offer scalable, customizable solutions that can be adapted to the diverse needs of people with dementia while minimizing mobility challenges [14,18,22-25].

Despite these promising findings, questions remain regarding the feasibility and the limitations of integrating VR into RT for dementia care. While several systematic and scoping reviews have examined the use of virtual reality reminiscence therapy (VRRT) for older adults or people with cognitive impairments [26,27], none, to the best of our knowledge, have focused exclusively on people with dementia. Moreover, existing reviews often conflate diverse populations, such as cognitively healthy older adults, those with mild cognitive impairment, and individuals with dementia, limiting the specificity and relevance of their conclusions to people with dementia. Further on the above, these prior works also tend to overlook important dimensions such as the comparative use of semi-immersive virtual reality (SI-VR) versus fully immersive virtual reality (FI-VR) systems, the role of content personalization, and implementation barriers in real-world care. This review addresses these critical gaps by focusing solely on people with dementia and providing a comprehensive synthesis of empirical and experimental studies that evaluate both the therapeutic impact and practical challenges of VR-based RT. As VR technology continues to evolve, further research is needed to determine its optimal application, long-term effects, and best practices for integration in dementia care. This study examines the implementation of VR in RT for people with dementia residing in long-term care facilities, hospital environments, or community settings over the past 2 and a half decades (2000 - 2025). By synthesizing findings from empirical and experimental research, the review addresses critical research questions (RQs), including the following:

- RQ1. How feasible is integrating VR into RT for dementia care?
- RQ2. What are the outcomes of VRRT?
- RQ3. What limitations currently hinder the application of VR in RT?
- RQ4. What are the potential future directions for research and implementation of VR in RT?

Methods

Literature Review Strategy

The electronic databases Google Scholar, ACM Digital Library, IEEE Xplore, MEDLINE, and PubMed were searched in January 2025 using a combination of search terms designed to capture studies on the use of VR in RT for people with dementia. Three core concept clusters were applied: term A included "dementia," and term B included "reminiscence" OR "reminiscence therapy," and term C included "virtual reality." A filter was applied to include only studies that were published between 2000 to 2025. The reference lists of articles that met the eligibility criteria were further perused to identify additional studies that may fall within the scope of this review.

Inclusion and Exclusion Criteria

Studies eligible to be included in this review had to meet the following inclusion criteria: (1) human participants were involved, (2) the full article was written in English, and (3) papers studied VR used for RT in dementia. The exclusion criteria were (1) publications where the study of VR used for RT in dementia was not the primary aim of the study, (2) publications that were not original studies (ie, review articles, letters, medical hypotheses, etc), (3) publications that presented trials studying subjects with no dementia, (4) duplicate publications, (5) publications whose abstract was not accessible, and (6) publications whose full text could not be obtained.

Data Collection Process

Following the identification of eligible publications, all relevant data were collected using a structured coding scheme in an Excel (Microsoft Corp) file. The data collected included titles, sample size, type of dementia, instruments used, methodology, and

findings. Additional fields specific to studies involving VR included the type of VR system (nonimmersive, semi-immersive, or fully immersive), VR content or intervention, feasibility, limitations, and future directions.

Data Synthesis and Analysis

This study used aggregated data where possible, per the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1: PRISMA-ScR). Key domains extracted included feasibility, therapeutic outcomes (emotional, cognitive, and social), and implementation barriers, allowing for cross-study comparison and pattern identification.

Results

Search Results

This search strategy yielded the identification of 136 articles. Following the eligibility assessment, 121 articles were excluded. In total, 15 papers satisfied the inclusion criteria and were used for this review (Table 1). Figure 1 illustrates the study selection process.

Figure 1. Flowchart detailing the process used to identify and select the papers included in the analysis.

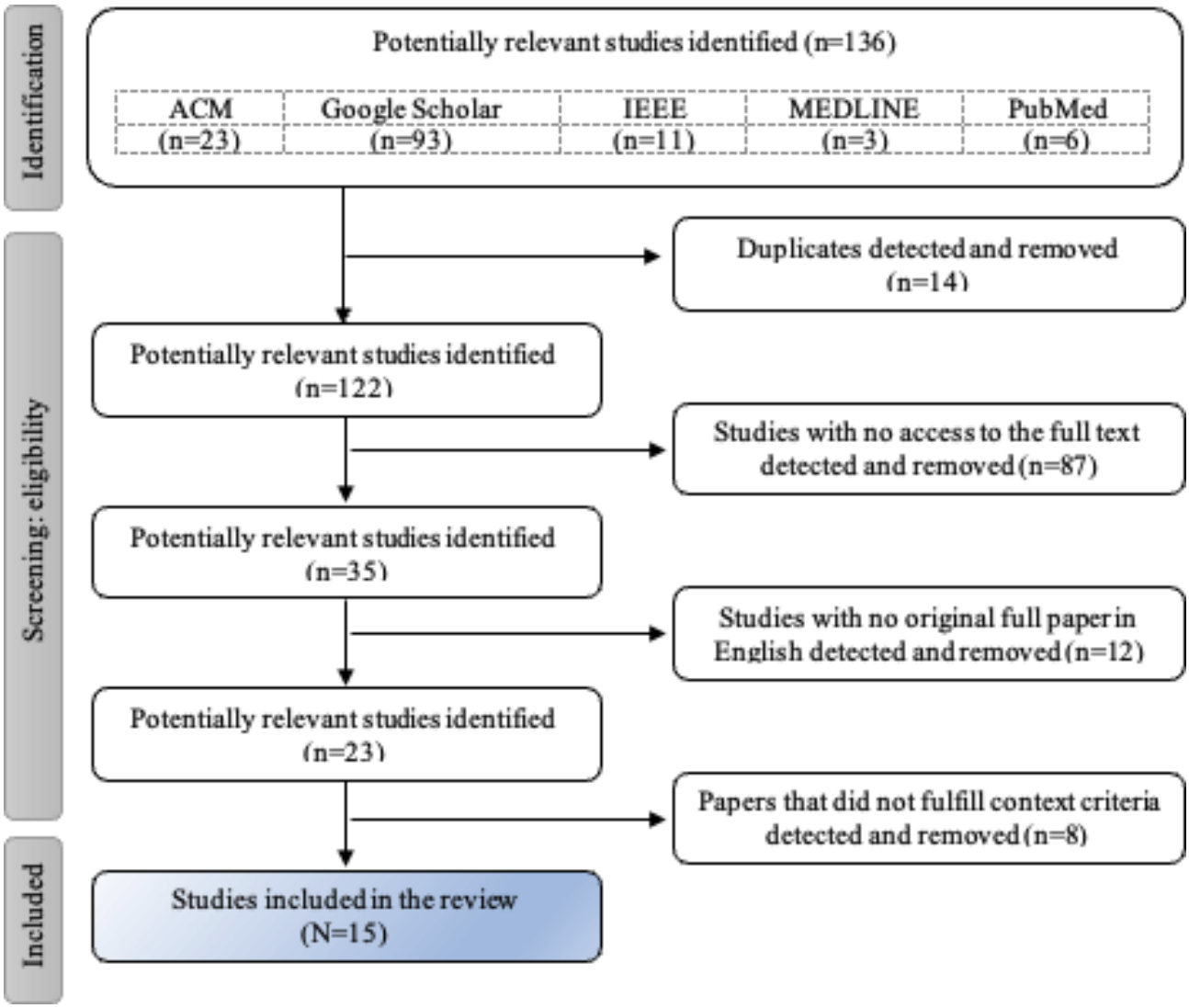


Table . Sample characteristics, dementia types, and severity levels.

	Study	Sample	Type of dementia	Level of dementia	Experimental design	RT[a] intervention
1	Appel et al, 2020 [16]	8 females/2 males, mean age: 86.5 years	AD[b], FD[c], MD[d], and VD[e]	Mild to severe	Mixed methods design, and combined quantitative and qualitative data	A single up to 20-minute FI-VR[f] RT session, focusing on managing BPSD[g]
2	Appel et al, 2024 [28]	45 females/24 males, age: 65+ years	N/A[h]	Mild to severe	Mixed methods design, randomized controlled trial, and combined quantitative and qualitative data	1 - 3 FI-VR sessions lasting up to 20 minutes each, focusing on managing BPSD
3	Brimelow et al, 2020 [29]	9 females/4 males, mean age: 66 - 93 years	N/A	Mild to moderate	Mixed methods design; combined quantitative observation and qualitative interviews post a single VR[i] session	A single up to 5-minute FI-VR RT individual or group session, focusing on apathy and mood
4	Coelho et al, 2020 [30]	6 females/3 males, mean age: 85.6 years	N/A	N/A	Mixed methods design, and combined quantitative and qualitative pre- and postdata	A total of 4 sessions lasting up to 15 minutes each of personalized FI-VR tailored to participants' psychological needs
5	Ferguson et al, 2020 [31]	22 females/3 males, mean age: 85 years	AD, MD, and VD	N/A	Mixed methods design, and combined quantitative and qualitative data	A single up to 30-minute FI-VR RT session, focusing on the feasibility of the system
6	Huang and Yang, 2022 [32]	11 females/9 males, mean age: 79 years	N/A	Mild to moderate	Longitudinal observational study design, and quantitative data	A 10 to 12-minute FI-VR RT session held twice a week for 3 months, targeting cognitive function, global status, and depression
7	Kim et al, 2021 [33]	10 females, mean age: 85.80 years	N/A	Mild	Mixed methods design, and combined quantitative and qualitative data, including a survey to gather information about the psychological needs of each patient, to customize the system	1 - 2 sessions lasting 20 - 30 minutes each of personalized FI-VR tailored to participants' psychological needs
8	Klein et al, 2018 [34]	3 females/3 males, mean age: 74.67 years	N/A	N/A	Qualitative methods design, and combined observational data and multiple focus groups before the experiment to gather information about the needs of people with dementia, to design the system	A single 12 - 20 minute FI-VR RT session, focusing on efficacy and acceptance of the system

	Study	Sample	Type of dementia	Level of dementia	Experimental design	RT[a] intervention
9	Manera et al, 2016 [35]	12 females/17 males, mean age: 76.3 years	AD and MD	N/A	Within-subject design, and combined quantitative and task performance data	A single 10-minute SI-VR[j] and controlled session, focusing on security, comfort, apathy, and anxiety
10	Moyle et al, 2018 [36]	7 females/3 males, mean age: 89 years	AD	N/A	Mixed methods design, and combined quantitative and qualitative data	A single up to 15-minute FI-VR RT session, focusing on engagement, apathy, and mood states
11	Ng et al, 2023 [17]	26 females/11 males, mean age: 65+ years	N/A	Mild	Nonrandomized controlled trial methods design and quantitative data	A single up to 12-minute FI-VR RT session, focusing on systems evaluation
12	Rose et al, 2021 [37]	2 females/6 males, mean age: 69.63 years	AD, FD, HD, and MD	Mild to severe	Mixed methods design, and quantitative observations and qualitative pre- and postdata	2 FI-VR RT sessions, lasting up to 15 minutes each, which focus on natural environments, aiming to reduce BPSD
13	Saredakis et al, 2020 [38]	10 females/7 males, mean age: 87.3 years	N/A	Mild to moderate	Mixed methods design and quantitative data	A single up to 20-minute FI-VR RT session, focusing on apathy
14	Saredakis et al, 2021 [39]	28 females/15 males, mean age: 84.8 years	N/A	N/A	Nonrandomized controlled trial methods design, and combined quantitative and objective data	A total of 3 FI-VR RT sessions, lasting up to 20 minutes each, focusing on apathy, cognition, depression, and quality of life
15	Tabbaa et al, 2019 [40]	2 females/6 males, mean age: 69.63 years	AD, MD, and VD	Mild to severe	Mixed methods design, and combined quantitative observations and qualitative pre- and postdata. A focus group was conducted before the experiment to gather information about the needs of people with dementia, to design the system	A total of 2 FI-VR RT sessions, lasting up to 15 minutes each, focusing on the technical aspects of the system, as well as its feasibility, acceptability, and practicality

[a]RT: reminiscence therapy.

[b]AD: Alzheimer disease.

[c]FD: frontotemporal dementia.

[d]MD: mixed dementia.

[e]VD: vascular dementia.

[f]FI-VR: fully immersive virtual reality.

[g]BPSD: behavioral and psychological symptoms of dementia.

[h]N/A: not available.

[i]VR: virtual reality.

[j]SI-VR: semi-immersive virtual reality.

[k]HD: Huntington disease.

Study Characteristics

The present review analyzes 15 studies, and the participant demographics include both male and female people with dementia, with mean ages ranging from 65 to 89 years. Several studies focus on specific types of dementia, including Alzheimer disease, vascular dementia, frontotemporal dementia, and mixed dementia, whereas others do not specify the type of dementia examined. The studies reflect a range of dementia severity from mild to severe, though the majority emphasize mild to moderate cases [17,27,38]. A summary of the demographics of the reviewed studies is presented in Table 1.

Types of VR Systems, Levels of Immersion, and Virtual Environments Used in Dementia RT

The reviewed studies implemented 2 levels of VR immersion, as detailed in Table 2, with FI-VR being the most commonly used method, present in 12 of 16 studies. FI-VR necessitated the use of head-mounted displays (HMDs), such as Samsung Gear VR, Oculus Rift, HTC VIVE, Oculus Quest, Oculus Go, and Windows Mixed Reality headsets, to provide 360-degree panoramic views and interactive experiences [27,33,39]. For instance, one study used the Samsung HMD Odyssey Windows Mixed Reality with Leap Motion sensors (LM-010), allowing people with dementia to interact with the VR environment using hand gestures, including controlling seasonal transitions and engaging with animated objects [33]. Similarly, 2 other studies used the HTC VIVE Pro to simulate a familiar home environment, incorporating daily life objects and home appliances [17,32]. Several studies used Samsung Gear VR with a smartphone, allowing people with dementia to immerse themselves in calming virtual environments, including underwater scenes, travel destinations, and snowy landscapes [27,37,40].

In 2 reviewed studies, the content was mirrored on an external flat screen, allowing caregiver participation in providing reassurance and guidance during VR sessions [37,40]. Additionally, 1 study created their custom HMD featuring 180-degree projection to simulate time travel through various historical periods, including Berlin and Paris in the 20th century [34].

SI-VR was used in 3 studies where people with dementia experienced virtual environments on large projection screens rather than HMDs [35,36]. This approach allowed for limited interaction through a mouse, touchscreen, or sensor-based technology, while still providing an engaging experience. For example, 1 study used stereoscopic 3D screens paired with Volfoni Edge 1.2 active 3D LCD shutter glasses to deliver an SI-VR experience [35]. Lastly, 2 additional studies incorporated SI-VR, projecting the virtual forest onto a large display screen, with sensor-based interactions that allowed people with dementia to explore the river, trees, and surrounding environment [36] as well as familiar and unfamiliar environments [30,33]. Nonimmersive VR was not featured in any of the studies examined.

Table . Apparatus and virtual environments.

	Study	Type of VR[a]	Virtual environment	Equipment
1	Appel et al, 2020 [16]	FI-VR[b]	Rocky lakeshore, forests, floating icebergs, and beaches	Samsung Gear VR and Sennheiser HD[c] 221 headphones
2	Appel et al, 2024 [28]	FI-VR	Calming in distinctive ways and nature visualization	Oculus Go with built-in and external Sennheiser HD 221 headphones
3	Brimelow et al, 2020 [29]	FI-VR	Underwater themes, beaches, farmyard animals, travel destinations, and snowscapes	Samsung Galaxy S7 and Samsung Gear VR headset
4	Coelho et al, 2020 [30]	FI-VR	Forests, beaches, cathedrals, childhood homes, workplaces, and religious venues	Samsung Gear VR and Oculus Rift VR
5	Ferguson et al, 2020 [31]	FI-VR	Beach scene	Mirage Solo with Daydream Business Edition
6	Huang and Yang, 2022 [32]	FI-VR	1960 - 1980 Taiwan: historical residence, radio, photo album, and feeding chickens	VIVE Pro
7	Kim et al, 2021 [33]	FI-VR	Streets of Memory, Nostalgic Youth, Homely Hometown, and Where I Want to Go	Leap Motion sensors and Samsung Odyssey Windows Mixed Reality Headset
8	Klein et al, 2018 [34]	Between SI-VR[d] and FI-VR	Time travel: Berlin (1970 - 1949), movie stars (1950s-1960s), television shows, and Paris in the 20th century	Custom-built HMD[e] (180-degree projection)
9	Manera et al, 2016 [35]	SI-VR	People	Barco OverView OLSF-721[f] full HD 3D stereoscopic LED video wall, Volfoni Edge 1.2 active 3D LCD shutter glasses
10	Moyle et al, 2018 [36]	SI-VR	Forest	Large screen
11	Ng et al, 2023 [17]	FI-VR	Early home environment: appliances and daily necessities	HTC Vive Pro and Leap Motion sensor
12	Rose et al, 2021 [37]	FI-VR	Nature and urban, forest, countryside, sandy or rocky beaches, and a cathedral	Samsung Gear VR and Samsung Galaxy S6
13	Saredakis et al, 2020 [38]	FI-VR	Personalized VR videos and places via Google Street View	Oculus Go
14	Saredakis et al, 2021 [39]	FI-VR	Personalized VR videos and places via Google Street View	Oculus Quest
15	Tabbaa et al, 2019 [40]	FI-VR	Cathedral, forest, sandy beach, rocky beach, and countryside	Samsung Gear VR and Samsung Galaxy S6 mobile phone

[a]VR: virtual reality.

[b]FI-VR: fully immersive virtual reality.

[c]HD: high definition.

[d]SI-VR: semi-immersive virtual reality.

[e]HMD: head-mounted display.

[f]OLSF: OverView LED Slim Front access.

VRRT and the Role of Nature, Nostalgic Memories, and Social Themes

Overall, our review findings indicate that the virtual environments designed for RT were mainly created to be calming and engaging, with a strong emphasis on nature, nostalgic memories, and social themes (Table 2). RT via VR aims to create immersive, emotionally meaningful experiences that promote relaxation and memory recall and enhance engagement for people with dementia. Numerous studies demonstrated that virtual environments featuring peaceful natural landscapes, historical experiences, and familiar settings contribute significantly to the emotional and psychological well-being of people with dementia [16,17,27,31,32,34,36,38].

One of the most widely used themes in RT via VR was nature, as research indicates that natural landscapes help induce relaxation and reduce agitation in people with dementia [16,17,27,36,37]. It was found that people with dementia who experienced virtual forests, lakes, and beaches reported reduced anxiety and enhanced emotional engagement [16]. It was further highlighted that participants displayed lower distress levels when immersed in nature-based VR experiences, particularly those incorporating sensory elements such as bird sounds, flowing water, and open landscapes [28,37].

Nostalgic memories were another central element in RT in VR, as many virtual environments replicate meaningful locations linked to the people with dementia's past that can trigger autobiographical memories [27,31,32,34,38]. People with dementia responded positively to virtual environments that resembled places from their younger years, leading to enhanced memory recall and more profound personal engagement with caregivers and peers. These places included simulated childhood homes, nostalgic town streets, familiar workplaces, and meaningful travel destinations, including time travel experiences [27,30,33,37,40]. For example, people with dementia were immersed in "Streets of Memory," a VR representation of old neighborhoods and markets [33], past workplaces such as farms and factories [30], cathedrals and holy places [40], natural landscapes such as beaches and forests that evoked memories of youth [27,31,37,40], and rural settings from 1950 - 1970 [32,34].

Finally, it was observed that people with dementia were more likely to participate in discussions and share personal experiences when the virtual environments replicated familiar social places [30,31,33,37,40]. These places included, for instance, a VR Christmas dinner [31], a cathedral church service [37], a retired teacher who experienced a VR classroom [30], a familiar-looking café [33], and a personalized Google street view experience, where people with dementia could visit places they remember, such as old neighborhoods, vacation spots, and religious sites [38,39].

Feasibility and Impact of VRRT

Overall, the reviewed studies consistently demonstrated that VR for RT is feasible and can have a positive impact on enhancing emotional well-being and social engagement in people with dementia (Table 3). Several studies confirmed that people with dementia could successfully engage with VR without experiencing significant adverse effects [27,28,30,33,40]. However, minor side effects, including dizziness, nausea, and discomfort, were reported [27,33,38], though these were temporary and did not affect the overall feasibility.

Further to the above, several studies reported that people with dementia were able to engage with VR environments, often requiring only initial guidance or passive supervision. While explicit data on caregiver workload was limited, these findings suggest the potential for VRRT to be implemented with manageable facilitation demands in structured care settings [27,28,30,40].

Moreover, the flexibility of VR delivery formats, specifically the availability of both fully immersive and semi-immersive systems, emerged as a significant facilitator of feasibility. Fully immersive systems, such as HMDs, provide a deeper sense of presence and sensory engagement, which can enhance therapeutic outcomes. However, these systems often come with higher costs and setup requirements. In contrast, semi-immersive systems using large screens or projection displays offer a more accessible and logistically manageable alternative, particularly beneficial for individuals with mobility limitations or in resource-constrained care settings. This dual-modality approach increases the adaptability of VRRT, making it feasible across a range of environments, from long-term residential facilities to community-based programs [27,30,38,40]. In summary, the findings from multiple studies indicate that VRRT is feasible and well-tolerated, with adaptable delivery formats that are suitable for various care settings.

Table . Reported side effects, feasibility, and outcomes of reminiscence therapy via virtual reality.

	Study	Instruments	Reported side effects	Feasibility or tolerance	Results
1	Appel et al, 2020 [16]	NPI[a], Confusion Assessment Method score, Montreal Cognitive Assessment, MMSE[b], recording instances of BPSD[c], and semistructured interviews.	Temporary feelings of dizziness and nausea.	People with dementia tolerated VR[d] well. VR is a feasible nonpharmacological intervention in acute care hospitals.	VR is a deployable, scalable, nonpharmacological solution for managing BPSD, which can significantly help dementia patients and their caregivers.
2	Appel et al, 2024 [28]	Quality of Life in Late-Stage Dementia Scale, nurses' daily notes for BPSDs and falls, and structured observations and interviews.	Two people with dementia experienced nervousness, anxiety, confusion, or disorientation, and 1 person with dementia experienced nausea.	People with dementia tolerated VR well. VR is a feasible nonpharmacological intervention in acute care hospitals.	VR is a safe, well-tolerated, and enjoyable nonpharmacological solution that can facilitate RT and significantly reduce aggressiveness in people with dementia.
3	Brimelow et al, 2020 [29]	PEAR[e], OERS[f], and structured observations and interviews.	Two people with dementia and impaired vision reported symptoms of cybersickness. One person with dementia found the headset slightly uncomfortable. Another person with dementia reported feeling "giddy," which was temporary upon device removal.	Mobile-based VR is feasible. People with dementia found VR enjoyable with low levels of physical and emotional discomfort.	The study found no impact on OERS measures; no significant increase in fear or anxiety. Reminiscence was observed in 6 of the 9 verbally communicative residents.
4	Coelho et al, 2020 [30]	Disability in daily activities (Barthel Index and Lawton and Brody Scale), Montreal Cognitive Assessment, Global Deterioration Scale, Cornell Scale for Depression in Dementia, NPI, SSQ[g], and EUROHIS-QOL-8.	N/A[h]	A feasible solution with no significant adverse effects related to simulator sickness or psychological and behavioral symptoms.	RT[i] via VR can benefit people with dementia, who are actively engaged in the sessions and share memories. No significant psychological or behavioral symptom changes were found.
5	Ferguson et al, 2020 [31]	Functional Assessment Staging Scale, PAINAD[j], and semistructured interviews.	Two people with dementia experienced worsened BPSD after VR exposure. Their PAINAD scores increased, indicating discomfort, distress, or pain.	VR is safe and enjoyable.	VR provides meaningful activity and enhances the quality of life for people with dementia.
6	Huang and Yang, 2022 [32]	Cognitive Abilities Screening Instrument, MMSE, Global status by Clinical Dementia Rating, and Depressive symptoms by the Center for Epidemiological Studies of Depression.	N/A	Feasible and well-tolerated.	RT via VR can improve mood and help preserve cognitive function in people with dementia during the intervention period.
7	Kim et al, 2021 [33]	MMSE, Activities of Daily Life, VR immersion scale, and observations.	Two people with dementia reported dizziness or nausea during VR exposure.	VR therapy was feasible and provided high satisfaction and immersion.	VR can be used to treat BPSD.

	Study	Instruments	Reported side effects	Feasibility or tolerance	Results
8	Klein et al, 2018 [34]	Semistructured interviews and observations.	N/A	Feasible when special consideration is given to choosing personally relevant and engaging content, as well as the therapy's contextual factors.	VR can enrich traditional RT, foster conversations, and support positive interactions between caregivers and people with dementia.
9	Manera et al, 2016 [35]	MMSE, Clinical Dementia Rating Scale-Sum of Box scores, diagnostic criteria for apathy in clinical practice, and attention tasks.	Low levels of discomfort, anxiety, and fatigue.	People with dementia reported high satisfaction and security.	VR can have a positive impact on people with dementia experiencing apathy.
10	Moyle et al, 2018 [36]	OERS, PEAR, semistructured interviews, and structured observations.	Some participants experienced mild fear or anxiety during VR.	Feasible and well-tolerated.	VR was perceived to have a positive effect on people with dementia. However, compared to the normative sample, a greater level of fear or anxiety was observed during VR. It may have the potential to improve quality of life.
11	Ng et al, 2023 [17]	N/A	Dizziness as an effect of VR teleporting.	Feasible but complex.	The study supports the use of VR for RT in people with dementia.
12	Rose et al, 2021 [37]	OERS, OAS-MNR[k], St Andrews Sexual Behavior Assessment, time exposed, and semistructured interviews.	One person with dementia reported dizziness due to the frequent movement of the headset to and from their eyes.	The study provides evidence of the clinical feasibility of VR implementation in health care settings.	VR can enhance the emotional well-being of people with dementia.
13	Saredakis et al, 2020 [38]	Psychogeriatric Assessment Scale, AES[l], SSQ, Slater-Usoh-Steed Presence Questionnaire, Phonemic and Semantic Verbal Fluency Tasks, expectations or enjoyment measure, and structured interview.	A total of 35% (6/17) of participants experienced temporary side effects such as discomfort around the cheekbone, nausea, and dizziness.	RT via VR is highly feasible.	People with dementia showed improved semantic scores immediately after using VR for RT. Those with higher levels of apathy demonstrated the greatest cognitive improvements after VR-RT[m].
14	Saredakis et al, 2021 [39]	AES, Addenbrooke Cognitive Examination III, Geriatric Depression Scale, Quality of Life in Alzheimer Disease, Three-Item Loneliness Scale, SSQ, and structured observations and interviews.	Two people with dementia reported aftereffects (headache and head feeling heavy) that occurred in the evening following a morning VR session.	VR can be implemented in an aged care setting with appropriate protocols in place.	People with dementia enjoyed RT via VR.

	Study	Instruments	Reported side effects	Feasibility or tolerance	Results
15	Tabbaa et al, 2019 [40]	OERS, OAS-MNR, semistructured interviews, and observational notes.	N/A	Feasible and well-tolerated.	VR enhanced the emotional well-being of people with dementia, with effects lasting for a short time after the session. VR also facilitated emotional openness between caregivers and people with dementia.

[a]NPI: Neuropsychiatric Inventory.

[b]MMSE: Mini-Mental State Exam.

[c]BPSD: behavioral and psychological symptoms of dementia.

[d]VR: virtual reality.

[e]PEAR: Person-Environment Apathy Rating.

[f]OERS: Observed Emotion Rating Scale.

[g]SSQ: Simulator Sickness Questionnaire.

[h]N/A: not available.

[i]RT: reminiscence therapy.

[j]PAINAD: Pain Assessment in Advanced Dementia.

[k]OAS-MNR: Overt Aggression Scale-Modified for Neurorehabilitation.

[l]AES: Apathy Evaluation Scale.

[m]VRRT: virtual reality reminiscence therapy.

VRRT and Its Impact on Emotional Well-Being

Emotional well-being is a vital aspect of dementia care, as people with dementia often face mood disturbances, anxiety, depression, and apathy. These issues are collectively known as the BPSD and can severely affect the quality of life of people with dementia [14,38]. Our review indicates that RT delivered through VR has been shown to have a positive impact on the BPSD, with numerous studies highlighting these benefits. Specifically, studies reported that the use of VR improved mood, reduced agitation, and increased engagement among people with dementia [27,31,35-38]. Additionally, RT via VR has been shown to alleviate emotional distress and psychological discomfort for both people with dementia and their caregivers, who were found to experience significant emotional relief and relaxation during VR reminiscence sessions [31,40], suggesting that VR can serve as a stress-reducing intervention.

One of the primary reasons RT via VR can enhance emotional well-being is its ability to immerse people with dementia in calming, familiar, or personally meaningful environments. Specifically, studies found that people with dementia who experienced VR environments depicting natural landscapes, such as forests, lakes, and beaches, exhibited increased relaxation and reduced anxiety [16,17,36]. Similarly, people with dementia who engaged in VR sessions featuring peaceful, scenic locations such as underwater themes, farmyards, or travel destinations displayed fewer signs of agitation and distress [27,34,38].

Furthermore, as with traditional RT, RT via VR can evoke deeply personal and emotionally meaningful experiences, which contribute to a sense of identity, self-awareness, and emotional fulfillment. It was observed that during RT in VR, people with dementia were able to restore memories from their past, triggering emotions tied to nostalgia, love, and belonging [39,40]. This emotional re-engagement often strengthens feelings of self-worth and dignity, which are crucial for maintaining well-being in dementia care [27]. To further support the above, another study found that VR reminiscence experiences elicited high levels of emotional satisfaction and security, with participants displaying more expressions of happiness and comfort compared to non-VR RT [35,37].

VRRT and Its Impact on Social Engagement

Social engagement is a crucial aspect of well-being for people with dementia, as it helps reduce loneliness, improve emotional stability, and strengthen relationships with caregivers and family members [38,41]. RT via VR has been shown to facilitate meaningful interactions by encouraging people with dementia to share personal experiences, engage in conversations, and participate in immersive social settings. Studies have consistently highlighted the ability of RT via VR to stimulate both verbal and nonverbal interactions, resulting in enhanced social connections [31,35,36,40]. To further support this, multiple studies observed that people with dementia who participated in VR reminiscence sessions were more likely to express emotions, initiate conversations, and reflect on past experiences. This enabled caregivers to gain deeper insights into their personal stories [39,40]. Additionally, research indicated that RT via VR enabled group discussions and storytelling, which encouraged people with dementia to comment on each other's experiences and engage in collective reminiscing, which reinforced social bonds [28,34,35].

Beyond verbal interactions, RT through VR also boosts emotional engagement, which plays a crucial role in maintaining social relationships [42]. Our review found that people with dementia who participated in RT via VR showed increased

smiling, laughter, and eye contact, indicating greater emotional connectivity with those around them [36,39]. These findings imply that VRRT not only triggers memories but also fosters present-moment social interactions that contribute to emotional well-being.

Another significant way RT via VR enhances social engagement is by reducing social withdrawal and apathy, common symptoms in dementia that often lead to isolation [38]. A study found that people with dementia who participated in RT via VR exhibited lower levels of social withdrawal and depression, indicating that immersive experiences can encourage active participation in social interactions [27,37]. This aligns with findings from another study [40], which noted that VR-based RT provided a shared platform for communication, making it easier for caregivers and people with dementia to connect over mutual experiences.

VRRT and Its Impact on Cognitive Stimulation

While RT via VR has shown promising effects on emotional well-being and social engagement, its impact on cognitive function remains inconclusive. In particular, a study suggests that RT via VR may help preserve cognitive function during the intervention period, but its long-term effects are unclear, requiring further research to determine whether these benefits persist over time [32].

On the other hand, several studies found no significant improvements in cognitive domains such as attention, memory retention, or processing speed, nor sufficient evidence that VR-based RT reduces cognitive decline over time [27,35,37]. However, people with dementia with higher levels of apathy exhibited significant cognitive improvements following RT via VR, suggesting that the therapy may be particularly beneficial for specific subgroups [38,39].

One possible explanation for these findings is that RT via VR primarily stimulates emotionally charged autobiographical memories, rather than engaging higher-order cognitive processes such as reasoning, problem-solving, or working memory [35,39]. While RT is known to activate episodic memory networks, the extent to which these activations contribute to broader cognitive function remains uncertain [27,39].

Additionally, RT via VR was found to elicit autobiographical memories in familiar environments, reinforcing the importance of personalized content [30,33]. This suggests that the success of RT via VR may depend on familiarity with the virtual environment, with well-known settings enhancing reminiscence and memory recall more successfully than unfamiliar ones.

Current Limitations and Future Directions in VRRT

Despite the promising potential of using VR to enhance RT, several limitations hinder its widespread implementation in dementia care settings. One of the primary challenges identified in the findings is VR technology's high cost and resource-intensive nature. Many long-term care facilities lack the financial resources to invest in expensive VR headsets, high-quality software, and the necessary infrastructure for setup and maintenance [40]. While SI-VR solutions using large screens provide a more affordable alternative, they lack the same level of immersion and engagement as FI-VR [35,36]. Additionally, RT via VR requires dedicated space and structured session planning, which can be difficult for care facilities with limited resources and understaffed teams. To address these issues, research should focus on developing cost-effective VR solutions for RT, such as mobile-based VR apps and lightweight, affordable VR headsets that require minimal setup [36]. Additionally, open-source VR software and cloud-based VR platforms could reduce infrastructure costs and improve accessibility for lower-resource care settings [40]. Scalable VR solutions with portable, low-cost hardware and preprogrammed virtual environments could make VR more practical for widespread clinical use [37].

Another significant limitation is the technical complexity of VR systems and the need for caregiver training. Facilitating an RT session using VR systems involves specialized equipment that requires caregivers to be trained in setup, troubleshooting, and guiding people with dementia through VR experiences [27,40]. However, many caregivers report limited confidence in using technology, and the high turnover rates in dementia care settings make continuous training difficult [37,40]. To improve usability and adoption, VR developers should focus on creating simplified, user-friendly interfaces that caregivers can operate with minimal training [16]. Structured VR training modules for caregivers, including interactive tutorials and hands-on workshops, could further support the implementation of VR in dementia care [31]. Future research should explore the integration of voice-assisted navigation and automated session setup, allowing caregivers to facilitate VR for RT with minimal assistance [27].

The findings also indicate that people with dementia exhibit individual variability in their response to VR, with some experiencing sensory overload, disorientation, or fatigue, making it necessary to adjust VR exposure based on individual needs [27,28,31]. However, many VR systems in dementia care lack adaptive features that personalize the therapeutic content based on cognitive and sensory preferences. Thus, adaptive VR experiences that adjust based on people with dementia's emotional and cognitive responses could help tailor RT sessions via VR to individual needs, reducing the risk of sensory overload [33].

Additionally, there is limited research on the long-term effects of VR in RT and its ability to sustain therapeutic benefits over time. While multiple studies confirm that VR can improve RT's outcomes, there is insufficient evidence on whether it provides lasting improvements [27]. Most existing studies focus on short-term effects, with limited follow-up on how repeated exposure to VR influences neurocognitive resilience in people with dementia. Furthermore, the effectiveness of VRRT compared to traditional RT remains unclear, raising questions about its long-term clinical value and cost-effectiveness [40]. Future research should prioritize longitudinal studies that assess the sustained effects of VR in RT over time [35]. Lastly, comparative studies evaluating VRRT against traditional RT could provide deeper insights into its long-term therapeutic potential and clinical relevance [36,40].

Discussion

Principal Findings

The findings of this review suggest that RT via VR is feasible and can positively enhance intervention in dementia care, with notable benefits in emotional well-being and social engagement. In particular, and in response to RQ1, it was found to be a feasible and well-tolerated intervention for people with dementia. Studies consistently report that people with dementia can successfully engage with virtual environments without experiencing major adverse effects. The availability of both fully immersive and semi-immersive systems enhances its adaptability, making it accessible for people with dementia with mild, moderate, and severe degrees of cognitive impairment and mobility limitations.

RT via VR has also been shown to enhance the health-related quality of life for people with dementia. In response to RQ2, the findings indicate that RT via VR addresses the BPSD by reducing agitation, anxiety, and apathy, while fostering relaxation and positive emotions, especially when immersed in peaceful, familiar, and personally meaningful VR environments. It is worth mentioning that the impact of RT via VR on cognition remains inconclusive. Studies examining its effects on memory recall, attention, and executive function found no significant improvements in cognitive performance. While RT via VR is able to stimulate autobiographical memory retrieval, it does not necessarily enhance long-term cognitive function or slow disease progression.

Comparison to Prior Work

The findings of this review are broadly consistent with recent reviews, both systematic and scoping, evaluating VRRT among older adults, including those with cognitive impairment. A systematic review [43] reported that VRRT is associated with emotional benefits such as reductions in anxiety, apathy, and depressive symptoms. Similarly, scoping reviews [26,27] found that immersive, autobiographically meaningful content tends to enhance mood and emotional engagement. However, this review also makes several distinct contributions to the literature. First, it focuses exclusively on people with dementia, while prior reviews included broader populations, such as older adults with or without cognitive impairment [26,27]. This more specific scope offers a dementia-targeted synthesis, enabling better applicability for practitioners and researchers in dementia care. Second, this review provides a more context-sensitive analysis of implementation challenges, extending beyond general usability concerns, identified in previous reviews such as simulator sickness, interface complexity, and the need for caregiver assistance [26,43]. This review builds upon prior usability discussions by identifying additional practical barriers that influence VRRT implementation, including the need for trained personnel, space constraints, and technical support requirements. While earlier reviews report general challenges, this review highlights how these factors can vary across care settings, suggesting the importance of context-specific implementation planning.

The cognitive outcomes reported remain limited across all reviews. While some evidence suggests that VRRT may stimulate autobiographical memory in the short term, none of the included reviews, including this one, found robust support for long-term improvements in memory, attention, or executive functioning. In previous reviews, the absence of randomized controlled trials to evaluate the solution has been emphasized [26,43], along with the need for more rigorous data [27]. Building on these insights, this review advocates for designing future studies around unified cognitive metrics and sustained postintervention follow-ups.

Lastly, this work offers forward-looking insights by providing practical, design-oriented recommendations, such as integrating mobile VR platforms, simplifying user interfaces, and creating culturally adaptive content libraries. These suggestions are aimed at addressing cost, accessibility, and caregiver usability gaps, areas not thoroughly operationalized in previous reviews [26,27]. As such, this review serves not only as a synthesis of existing evidence but also as a strategic roadmap for designing more scalable and feasible VRRT interventions explicitly tailored for dementia care.

Strengths and Limitations

This review provides a comprehensive synthesis of current evidence on the feasibility, acceptability, and therapeutic potential of delivering RT through VR for people with dementia. By incorporating studies that span varying levels of cognitive impairment and care environments, it offers a nuanced understanding of how both SI-VR and FI-VR technologies can be adapted to meet the complex and evolving needs of this population. A key contribution of this review lies in its focused analysis of critical dimensions often overlooked in prior reviews, including the degree of immersion, personalization, content relevance, and implementation challenges within real-world care settings.

To the best of our knowledge, this is the only recent review that concentrates solely on people with dementia, excluding broader populations of older adults with or without cognitive impairment, while examining the differential use of SI-VR and FI-VR systems. Moreover, by emphasizing system design considerations, the review aligns with contemporary calls for more inclusive, accessible, and person-centered VR interventions. These insights extend the review's relevance beyond clinical practitioners to include gerontologists, human-computer interaction researchers, and technology developers working at the intersection of dementia care and digital innovation. In doing so, it contributes a timely and interdisciplinary perspective that can guide the design, deployment, and evaluation of future VR-based therapeutic tools.

However, in response to RQ3, several limitations must be acknowledged. First, despite the promising findings, the high cost of VR equipment, software, and required infrastructure remains a significant barrier to widespread adoption, especially in resource-limited care settings. While semi-immersive systems may offer more accessible alternatives, they often lack the engagement and immersive quality necessary to maximize therapeutic effects. Second, the technical complexity of VR systems presents an operational challenge. Caregivers must receive adequate training to operate equipment and support

people with dementia through virtual experiences, which may be unrealistic in understaffed environments.

Furthermore, many of the included studies are limited by small sample sizes, short intervention durations, and a lack of standardized outcome measures, which hinders cross-study comparability. Research on the long-term efficacy of VRRT is particularly scarce, with most studies focusing on immediate or short-term outcomes. The variability in individual responses also complicates implementation; while some people with dementia benefit substantially, others may experience sensory overload, fatigue, or disengagement. Current VR platforms generally lack adaptive features that can tailor the experience to individual user profiles or adjust dynamically in response to behavioral feedback.

In addition to the challenges identified in response to RQ3, this review itself is subject to several methodological limitations. First, the literature search was limited to freely accessible full-text papers published in English. As a result, relevant studies published behind paywalls or in other languages may have been excluded, potentially narrowing the comprehensiveness and representativeness of the findings. Second, the review was conducted by a single author, who carried out all stages of the process, including study screening, selection, and data extraction. While efforts were made to apply clear and consistent criteria throughout, the absence of a second reviewer may have introduced potential selection bias and reduced intercoder reliability.

Future Directions

To address RQ4, future research should enhance affordability, personalization, long-term impact, and usability of VRRT for people with dementia. Developing cost-effective, mobile-based VR solutions, such as lightweight headsets or tablet-compatible apps, could significantly enhance accessibility, particularly in underresourced care settings. These alternatives should aim to retain therapeutic immersion while reducing financial and infrastructural burdens associated with high-end VR systems.

Moreover, the personalization of VRRT experiences remains critical to their effectiveness. Future systems should integrate adaptive features that tailor content based on individual life histories, preferences, and cognitive or sensory needs. Approaches such as user-driven content selection, biometric feedback integration, and modular content frameworks can help dynamically adjust the experience to match user tolerance and emotional state.

To maximize real-world applicability, simplified interfaces and structured caregiver training protocols are necessary to reduce technical complexity and empower care staff. The inclusion of training toolkits, step-by-step onboarding modules, and in-app guidance systems could support smoother integration in everyday care workflows.

Importantly, research should also focus on establishing the long-term effects of VRRT through high-quality, longitudinal studies. These should investigate not only sustained psychological and cognitive outcomes but also the potential for reducing caregiver burden and improving overall health-related quality of life. Comparative trials between VRRT and traditional RT are also needed to clarify cost-effectiveness and relative efficacy.

Finally, co-design methodologies that involve people with dementia, caregivers, and interdisciplinary experts in the development process can ensure that future VRRT systems are ethically grounded, emotionally safe, and attuned to the lived realities of diverse dementia care populations.

Conclusions

In conclusion, this scoping review illustrates that the employment of RT through VR provides a viable and innovative strategy for dementia care. Nevertheless, practical challenges such as considerable costs, the complexity of personalization, and the necessity for caregiver training impede its extensive implementation within care environments. Notwithstanding these obstacles, the findings highlight the potential of VRRT to function as a scalable, person-centered tool in dementia care. Subsequent research that addresses usability, cost-effectiveness, and long-term outcomes, while integrating inclusive co-design methodologies, can contribute to the evolution of VRRT into a sustainable and equitable element in dementia care.

Data Availability

All data generated or analyzed during this study are included in this published paper (and its supplementary information files).

Authors' Contributions

MM conceived and designed this work, performed its data analysis and interpretation, and drafted the paper.

Conflicts of Interest

MM is an associate editor for *JMIR XR and Spatial Computing* at the time of this publication.

Checklist 1
PRISMA-ScR. PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews.
[DOCX File, 41 KB - xr_v2i1e73539_app1.docx]

References

1. Dementia fact sheet. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2025-07-09]
2. Kales HC, Gitlin LN, Lyketsos CG. Assessment and management of behavioral and psychological symptoms of dementia. BMJ 2015 Mar 2;350:h369. [doi: 10.1136/bmj.h369] [Medline: 25731881]
3. White N, Leurent B, Lord K, Scott S, Jones L, Sampson EL. The management of behavioural and psychological symptoms of dementia in the acute general medical hospital: a longitudinal cohort study. Int J Geriatr Psychiatry 2017 Mar;32(3):297-305. [doi: 10.1002/gps.4463] [Medline: 27019375]
4. Banerjee S. The Use of Antipsychotic Medication for People with Dementia: Time for Action: Department of Health; 2009.
5. Banerjee S, Hellier J, Dewey M, et al. Sertraline or mirtazapine for depression in dementia (HTA-SADD): a randomised, multicentre, double-blind, placebo-controlled trial. The Lancet 2011 Jul;378(9789):403-411. [doi: 10.1016/S0140-6736(11)60830-1] [Medline: 21371651]
6. Woods RT, Orrell M, Bruce E, et al. REMCARE: pragmatic multi-centre randomised trial of reminiscence groups for people with dementia and their family carers: effectiveness and economic analysis. PLoS ONE 2016;11(4):e0152843. [doi: 10.1371/journal.pone.0152843] [Medline: 27093052]
7. Kasl-Godley J, Gatz M. Psychosocial interventions for individuals with dementia: an integration of theory, therapy, and a clinical understanding of dementia. Clin Psychol Rev 2000 Aug;20(6):755-782. [doi: 10.1016/s0272-7358(99)00062-8] [Medline: 10983267]
8. Edmeads J, Metatla O. Designing for reminiscence with people with dementia. Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290607.3313059]
9. Charlesworth G, Burnell K, Crellin N, et al. Peer support and reminiscence therapy for people with dementia and their family carers: a factorial pragmatic randomised trial. J Neurol Neurosurg Psychiatry 2016 Nov;87(11):1218-1228. [doi: 10.1136/jnnp-2016-313736] [Medline: 27521377]
10. Yu F, Mathiason MA, Johnson K, Gaugler JE, Klassen D. Memory matters in dementia: efficacy of a mobile reminiscing therapy app. Alzheimers Dement (N Y) 2019;5(1):644-651. [doi: 10.1016/j.trci.2019.09.002] [Medline: 31720365]
11. Moon S, Park K. The effect of digital reminiscence therapy on people with dementia: a pilot randomized controlled trial. BMC Geriatr 2020 May 6;20(1):166. [doi: 10.1186/s12877-020-01563-2] [Medline: 32375661]
12. Quail Z, Carter MM, Wei A, Li X. Management of cognitive decline in Alzheimer's disease using a non-pharmacological intervention program: a case report. Medicine (Baltimore) 2020 May 22;99(21):e20128. [doi: 10.1097/MD.0000000000020128] [Medline: 32481282]
13. Appel L, Ali S, Narag T, et al. Virtual reality to promote wellbeing in persons with dementia: a scoping review. J Rehabil Assist Technol Eng 2021;8:20556683211053952. [doi: 10.1177/20556683211053952] [Medline: 35024166]
14. Matsangidou M, Solomou T, Frangoudes F, Papayianni E, Kkeli N, Pattichi CS. "Transported to a better place": the influence of virtual reality on the behavioural and psychological symptoms of dementia. Int J Hum Comput Stud 2025 Feb;196:103413. [doi: 10.1016/j.ijhcs.2024.103413]
15. Rose V, Stewart I, Jenkins KG, Ang CS, Matsangidou M. A scoping review exploring the feasibility of virtual reality technology use with individuals living with dementia. ICAT-EGVE 2018:131-139. [doi: 10.2312/egve.20181325]
16. Appel L, Kisonas E, Appel E, et al. Introducing virtual reality therapy for inpatients with dementia admitted to an acute care hospital: learnings from a pilot to pave the way to a randomized controlled trial. Pilot Feasibility Stud 2020 Oct 31;6(1):166. [doi: 10.1186/s40814-020-00708-9] [Medline: 33292729]
17. Ng HF, Cheng AS, Tai KSK, et al. Reminiscence therapy in virtual reality: a virtual reality application with local content to help mild dementia. AJSoTL 2023;13(1):81-99 [FREE Full text]
18. Matsangidou M, Solomou T, Frangoudes F, Papayianni E, Pattichis CS. Offering outworld experiences to in-patients with dementia through virtual reality: mixed methods study. JMIR Aging 2023 Aug 31;6(1):e45799. [doi: 10.2196/45799] [Medline: 37656031]
19. Appel L, Kisonas E, Appel E, et al. Administering virtual reality therapy to manage behavioral and psychological symptoms in patients with dementia admitted to an acute care hospital: results of a pilot study. JMIR Form Res 2021 Feb 3;5(2):e22406. [doi: 10.2196/22406] [Medline: 33533720]
20. D'Cunha NM, Nguyen D, Naumovski N, et al. A mini-review of virtual reality-based interventions to promote well-being for people living with dementia and mild cognitive impairment. Gerontology 2019;65(4):430-440. [doi: 10.1159/000500040] [Medline: 31108489]
21. Optale G, Urgesi C, Busato V, et al. Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. Neurorehabil Neural Repair 2010 May;24(4):348-357. [doi: 10.1177/1545968309353328] [Medline: 19934445]
22. Matsangidou M, Frangoudes F, Hadjiaros M, et al. "Bring me sunshine, bring me (physical) strength": the case of dementia. Designing and implementing a virtual reality system for physical training during the COVID-19 pandemic. Int J Hum Comput Stud 2022 Sep;165:102840. [doi: 10.1016/j.ijhcs.2022.102840]
23. Matsangidou M, Frangoudes F, Schiza E, et al. Participatory design and evaluation of virtual reality physical rehabilitation for people living with dementia. Virtual Real 2023 Mar;27(1):421-438. [doi: 10.1007/s10055-022-00639-1]

24. Matsangidou M, Schiza E, Hadjiaros M, et al. Dementia: I am physically fading. Can virtual reality help? Physical training for people with dementia in confined mental health units. Presented at: International Conference on Human-Computer Interaction; Jul 19, 2020; Copenhagen, Denmark. [doi: 10.1007/978-3-030-49282-3_26]
25. Matsangidou M, Solomou T, Frangoudes F, et al. Affective out-world experience via virtual reality for older adults living with mild cognitive impairments or mild dementia. Int J Environ Res Public Health 2023 Feb 7;20(4):2919. [doi: 10.3390/ijerph20042919] [Medline: 36833616]
26. Mao Q, Zhao Z, Yu L, Zhao Y, Wang H. The effects of virtual reality-based reminiscence therapies for older adults with cognitive impairment: systematic review. J Med Internet Res 2024 Nov 12;26:e53348. [doi: 10.2196/53348] [Medline: 39531267]
27. Pardini S, Calcagno R, Genovese A, Salvadori E, Ibarra OM. Exploring virtual reality-based reminiscence therapy on cognitive and emotional well-being in people with cognitive impairments: a scoping review. Brain Sci 2025 May 13;15(5):500. [doi: 10.3390/brainsci15050500] [Medline: 40426671]
28. Appel L, Appel E, Kisonas E, et al. Evaluating the impact of virtual reality on the behavioral and psychological symptoms of dementia and quality of life of inpatients with dementia in acute care: randomized controlled trial (VRCT). J Med Internet Res 2024 Jan 30;26:e51758. [doi: 10.2196/51758] [Medline: 38289666]
29. Brimelow RE, Dawe B, Dissanayaka N. Preliminary research: virtual reality in residential aged care to reduce apathy and improve mood. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):165-170. [doi: 10.1089/cyber.2019.0286] [Medline: 31829729]
30. Coelho T, Marques C, Moreira D, et al. Promoting reminiscences with virtual reality headsets: a pilot study with people with dementia. Int J Environ Res Public Health 2020 Dec 12;17(24):1-13. [doi: 10.3390/ijerph17249301] [Medline: 33322679]
31. Ferguson C, Shade MY, Boron JB, Lyden E, Manley NA. Virtual reality for therapeutic recreation in dementia hospice care: a feasibility study. Am J Hosp Palliat Care 2020 Oct;37(10):809-815. [doi: 10.1177/1049909120901525] [Medline: 31975609]
32. Huang LC, Yang YH. The long-term effects of immersive virtual reality reminiscence in people with dementia: longitudinal observational study. JMIR Serious Games 2022 Jul 25;10(3):e36720. [doi: 10.2196/36720] [Medline: 35877169]
33. Kim JH, Park S, Lim H. Developing a virtual reality for people with dementia in nursing homes based on their psychological needs: a feasibility study. BMC Geriatr 2021 Mar 7;21(1):167. [doi: 10.1186/s12877-021-02125-w] [Medline: 33678160]
34. Klein P, Uhlig M, Will H. The touch and feel of the past—using haptic and VR artefacts to enrich reminiscence therapy for people with dementia. Technologies (Basel) 2018;6(4):104. [doi: 10.3390/technologies6040104]
35. Manera V, Chapoulie E, Bourgeois J, et al. A feasibility study with image-based rendered virtual reality in patients with mild cognitive impairment and dementia. PLoS ONE 2016;11(3):e0151487. [doi: 10.1371/journal.pone.0151487] [Medline: 26990298]
36. Moyle W, Jones C, Dwan T, Petrovich T. Effectiveness of a virtual reality forest on people with dementia: a mixed methods pilot study. Gerontologist 2018 May 8;58(3):478-487. [doi: 10.1093/geront/gnw270] [Medline: 28329867]
37. Rose V, Stewart I, Jenkins KG, Tabbaa L, Ang CS, Matsangidou M. Bringing the outside in: the feasibility of virtual reality with people with dementia in an inpatient psychiatric care setting. Dementia (London) 2021 Jan;20(1):106-129. [doi: 10.1177/1471301219868036] [Medline: 31510801]
38. Saredakis D, Keage HA, Corlis M, Loetscher T. Using virtual reality to improve apathy in residential aged care: mixed methods study. J Med Internet Res 2020 Jun 26;22(6):e17632. [doi: 10.2196/17632] [Medline: 32469314]
39. Saredakis D, Keage HA, Corlis M, Ghezzi ES, Loffler H, Loetscher T. The effect of reminiscence therapy using virtual reality on apathy in residential aged care: multisite nonrandomized controlled trial. J Med Internet Res 2021 Sep 20;23(9):e29210. [doi: 10.2196/29210] [Medline: 34542418]
40. Tabbaa L, Ang CS, Rose V, et al. Bring the outside in: providing accessible experiences through VR for PWD in locked psychiatric hospitals. Presented at: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Edinburgh, Scotland. [doi: 10.1145/3290605.3300466]
41. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. The Lancet 2017 Dec;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6]
42. Fazio S, Pace D, Flinner J, Kallmyer B. The fundamentals of person-centered care for individuals with dementia. Gerontologist 2018 Jan 18;58(suppl_1):S10-S19. [doi: 10.1093/geront/gnx122] [Medline: 29361064]
43. Ng WHD, Ang WHD, Fukahori H, et al. Virtual reality-based reminiscence therapy for older adults to improve psychological well-being and cognition: a systematic review. J Clin Nurs 2024 Jul 17;33(3–4):e12918. [doi: 10.1111/jocn.17375] [Medline: 39020507]

Abbreviations

BPSD: behavioral and psychological symptoms of dementia
FI-VR: fully immersive virtual reality
HMD: head-mounted display

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RQ: research question
RT: reminiscence therapy
SI-VR: semi-immersive virtual reality
VR: virtual reality
VRRT: virtual reality reminiscence therapy

Edited by T Leung; submitted 06.03.25; peer-reviewed by B Tosti, J Wong; revised version received 23.06.25; accepted 23.06.25; published 06.08.25.

Please cite as:
Matsangidou M
Virtual Reality Reminiscence Therapy in Dementia Care: Scoping Review of Research
JMIR XR Spatial Comput 2025;2:e73539
URL: https://xr.jmir.org/2025/1/e73539
doi:10.2196/73539

Applications of Augmented Reality for Prehospital Emergency Care: Systematic Review of Randomized Controlled Trials

Rayan E Harari[1,2,3], PhD; Sara L Schulwolf[4], BA; Paulo Borges[1], MD; Hamid Salmani[5], MSc; Farhang Hosseini[5], PhD; Shannon K T Bailey[6], PhD; Brian Quach[1], BSc; Eric Nohelty[1], BSc; Sandra Park[1], MD; Yash Verma[1], BSc; Eric Goralnick[1], MD, PhD; Scott A Goldberg[1], MD; Hamid Shokoohi[1], MD; Roger D Dias[1], MD, PhD; Andrew Eyre[1], MD

[1]Department of Emergency Medicine, Mass General Brigham, Harvard Medical School, 10 Vining Street, STRATUS Center for Medical Simulation, Boston, MA, United States

[2]Quantitative Musculoskeletal Imaging Group Research, Department of Radiology, Mass General Brigham, Boston, MA, United States

[3]Harvard Data Science Initiative, Harvard Medical School, Boston, MA, United States

[4]University of Connecticut School of Medicine, Farmington, CT, United States

[5]MetaLeaps Inc, Boston, MA, United States

[6]Center for Advanced Medical Learning & Simulation, University of South Florida, Tampa, FL, United States

Corresponding Author:
Rayan E Harari, PhD
Department of Emergency Medicine, Mass General Brigham, Harvard Medical School, 10 Vining Street, STRATUS Center for Medical Simulation, Boston, MA, United States

Abstract

Background: Delivering high-quality prehospital emergency care remains challenging, especially in resource-limited settings where real-time clinical decision support is limited. Augmented reality (AR) has emerged as a promising health care technology, offering potential solutions to enhance decision-making, care processes, and emergency medical service (EMS) training.

Objective: This systematic review assesses the effectiveness of AR in improving clinical decision-making, care delivery, and educational outcomes for EMS providers.

Methods: We searched databases including PubMed, Cochrane CENTRAL, Web of Science, Institute of Electrical and Electronics Engineers (IEEE), Embase, PsycInfo, and Association for Computing Machinery (ACM). Studies were selected based on their focus on AR in prehospital care. A total of 14 randomized controlled trials were selected from an initial screening of 2081 manuscripts. Included studies focused on AR use by EMS personnel, examining clinical and educational impacts. Data such as study demographics, intervention type, outcomes, and methodologies were extracted using a standardized form. Primary outcomes assessed included clinical task accuracy, response times, and training efficacy. A narrative synthesis was conducted, and bias was evaluated using Cochrane's risk of bias tool. Improvements in AR-assisted interventions and their limitations were analyzed.

Results: AR significantly improved clinical decision-making accuracy and EMS training outcomes, reducing response times in simulations and real-world applications. However, small sample sizes and challenges in integrating AR into workflows limit the generalizability of the findings.

Conclusions: AR holds promise for transforming prehospital care by enhancing real-time decision-making and EMS training. Future research should address technological integration and scalability to fully realize AR's potential in EMS.

(JMIR XR Spatial Comput 2025;2:e66222) doi:10.2196/66222

KEYWORDS

prehospital emergency care; emergency medical services; randomized controlled trials; clinical decision support; training; augmented reality; emergency; care; systematic review; BLS; procedures; traumatic injury; survival; prehospital; emergency care; AR; decision-making; educational; education; EMS; database; technology; critical care; basic life support

Introduction

Overview

The prehospital setting represents a critical area of emergency medical care. Emergency medical services (EMSs) providers, such as emergency medical technicians, firefighters, and paramedics care for diverse patient populations in variable in highly acute settings; they are often the first to respond to life-threatening scenarios such as traumatic injury or cardiac arrest. Innovations in prehospital care have led to improvement in patient outcomes over the past several decades, including a

reduction in early deaths following traumatic injuries and improved survival from out-of-hospital cardiac arrest following early initiation of basic life support (BLS) procedures [1-5]. However, there remain significant challenges to providing high-quality prehospital emergency care, especially in resource-limited settings. Prehospital emergency care literature reports that top research priorities include augmenting the education and training of EMS personnel as well as improving the management of patients with life-threatening conditions such as asthma exacerbation, traumatic brain injury, and cardiac ischemia [6,7]. Further, improving the availability and response quality of medical control physicians for EMS systems has been cited as an additional area of interest [8].

With the need for improvements in both real-time decision support in prehospital care and the education and training of prehospital care providers, researchers have posited the utility of integrating AR into the prehospital setting. AR technologies are tools to superimpose digitally generated 3D and 2D visual information into a user's environment in real time for display and guidance. Unlike virtual reality, in which a user is completely immersed in a virtual environment that occludes their physical environment, users of AR technologies can interact with both their physical environment and digitally generated images [9].

AR already has significant implications within health care, with AR-based clinical and training modalities beginning to emerge within several medical fields [10-13]. The most well-documented examples come from surgical specialties, which have for years used AR-based equipment as clinical decision support (CDS) and training tools to practice intricate procedures; additionally, many subdisciplines including bariatric surgery, oral-maxillofacial surgery, and neurosurgery use AR-based minimally-invasive robotic procedures [14-19]. Experts have suggested that AR-based CDS tools may prove useful to a variety of prehospital applications, such as providing real-time decision support for patient resuscitation or enhancing BLS education.

To date, there have been few systematic examinations of AR in emergency medicine (EM), with even fewer specifically investigating prehospital emergency medical care. This manuscript thus presents a systematic review of randomized control trials (RCTs) investigating applications of AR in prehospital emergency medical care. Our primary objective is to evaluate the efficacy and effectiveness of AR applications in improving patient outcomes, care processes, and learning outcomes in the prehospital emergency care setting. Our secondary objectives are to identify challenges and limitations for the implementation of AR-based CDS and training tools in prehospital EM and to explore future directions for AR applications in these domains.

Methods

Literature Search

A systematic review of the available literature was performed to investigate the effect of AR on prehospital emergency medical care. Eligibility criteria for inclusion in the systematic review included peer-reviewed manuscripts published between 1970 and 2024 (June 10) in English-language journals. A search was conducted of online academic databases including PubMed, CENTRAL, Web of Science, Institute of Electrical and Electronics Engineers (IEEE), Embase, PsycInfo, CINAHL Complete, and Association for Computing Machinery (ACM). Detailed search strategy across databases for identifying studies on AR in prehospital emergency care can be found in Multimedia Appendix 1.

Full-Text Review

A search of these 8 academic databases yielded 2081 manuscripts for review. Two independent reviewers first screened titles and abstracts to remove duplicates (n=726) as well as manuscripts that were not related to EM (n=1228). A full-text review of 127 studies was conducted by 8 independent researchers to assess their eligibility. Studies were included in full-text screening if a reviewer consensus of 2 reviewers deemed the study eligible. Each study during full-text screening was reviewed by 2 of the 8 reviewers independently and consensus was determined by a third reviewer. Data extraction was conducted independently by 2 reviewers using Covidence software (Veritas Health Innovation), which facilitated the management and review of manuscripts. Each reviewer independently extracted data, including study characteristics, participant demographics, intervention details, and outcome measures. Any discrepancies in the extracted data were resolved through discussion, with a third reviewer stepping in to make the final decision when necessary. No automation tools were used in the data extraction process. The full data extraction form can be seen in Multimedia Appendix 2.

Criteria for Inclusion

Criteria for inclusion into the final systematic review included full RCT or crossover RCT design; study setting in an EM; and use of wearable, handheld, or projection-based AR in intervention. Studies were included if they investigated the impact of AR on health care professionals or health care students, including emergency responders, paramedics, emergency medical technicians, medics, EM physicians, residents, or fellows, physician assistants, medical and health care students, surgeons, nurses, firefighters, law enforcement officers, or other relevant population (eg, lifeguards, other university students and lay first-responders, or unspecified medical specialties). Studies were also excluded if they were only a description of the technology without learning, performance, or other intervention outcomes.

Key Data Extracted

Primary outcomes of interest included patient outcomes or clinical performance outcomes such as task completion time, accuracy, number of attempts, and errors. Secondary outcomes included user experience or human factors outcomes such as technology acceptance, workload, stress, and cyber- or simulator-sickness. Key data for analysis was extracted from each of the included manuscripts by 2 independent reviewers using a standardized data extraction form. All data were collected and recorded using Microsoft Excel software. Data collected included study characteristics, participant

demographics, AR information, outcome measures, results, and limitations.

In addition to primary outcome measures such as task completion time, procedure accuracy, and protocol compliance, we collected data on several other key variables. These included study characteristics (publication year, country of study, design type, sample size), participant characteristics (professional roles such as first responders, paramedics, medical students; study population size; and whether the setting was civilian or military). Intervention characteristics were also documented, focusing on the type of AR platform used (eg, HoloLens, Vuzix, and Google Glasses) and the intervention context (real-time clinical support or educational training). Secondary outcome measures like user experience, technology acceptance, workload, and the occurrence of simulator sickness were also analyzed. No assumptions were made about missing or unclear data, and any such data were marked as "not reported."

Consensus

Consensus between reviewers was tracked via Microsoft Excel spreadsheet and calculated using Cohen κ, with an average of 0.71 (95% CI 0.635 - 0.785). The quality and potential bias of the included studies were evaluated on a manuscript level by independent reviewers using Cochrane's risk of bias tool [20], which can be seen in Multimedia Appendix 3, and reviewed by group consensus. The literature review and evaluation process are detailed in Figure 1. All data were summarized collectively and reported as an aggregate as well as in subgroups including "education and training" and "clinical decision making". Qualitative and descriptive data were synthesized narratively. The review protocol can be accessed in the Multimedia Appendix 2.

Results

Characteristics of Included Studies

Figure 1 presents the review procedure and the resulting number of relevant papers based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [21]. The characteristics of the 14 studies included in this systematic review are summarized in Table 1.

Figure 1. Systematic literature review procedure and the resulting number of relevant papers using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [21]. RCT: randomized controlled trial.

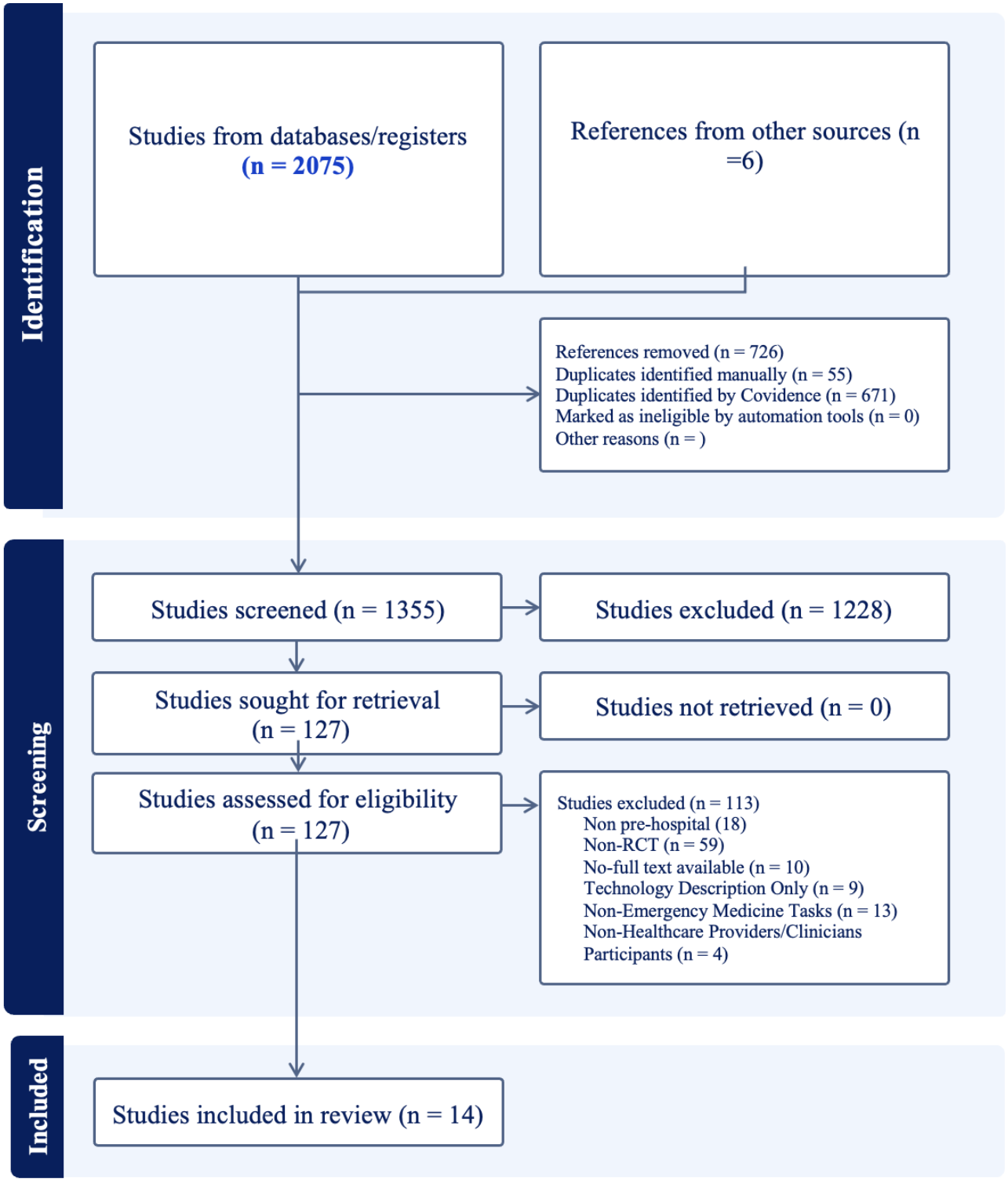


Table . Summary of studies evaluating augmented reality (AR) interventions in prehospital care, including study populations, AR platforms used, primary outcomes, and main findings across various emergency medical scenarios.

First author, publication year	Study population and sample size	AR intervention; platform	Primary outcome measures	Main findings
Rebol et al, 2023 [22]	First responders (n=25)	Real-time assistance for CPR[a] performance; HoloLens	CPR performance metrics (compression depth and rate)	No significant performance difference between mixed reality and control group
Koutitas et al, 2019 [23]	EMS[b] cadets (n=30)	Training module for the operation of AmBus systems; HoloLens	Time to task completion and error rate	Significant reduction in task completion time and error rate in AR group
Gruenerbl et al, 2018 [24]	Nursing students (n=50)	CPR training module; Google glasses	CPR performance metrics (compression depth and rate) before and after training	Significant improvement in posttraining performance in AR group
Doswell et al, 2020 [25]	First responders (n=10)	BLS[c] procedures training module; HoloLens	Time to correct procedure performance	No significant difference in performance time between AR and control group
Collington et al, 2018 [26]	Firefighters (n=10)	BLS procedures training module; Moverio glasses	Performance in simulated trauma scenarios	Significant improvement in self-reported hands-on skills proficiency in AR group
Barcala-Furelos et al, 2023 [27]	Lifeguards (n=38)	Real-time assistance for simulated infant delivery; Vuzix	Performance time and compliance with protocol	Significantly improved protocol adherence in AR group
Follman et al, 2019 [28]	Paramedics (n=31)	Real-time assistance in MCI[d] triage; ReconJet	Screening time and assessment accuracy	Significant improvement in triage accuracy in AR group
Du et al, 2022 [29]	Medical students (n=20)	Tactical Combat Casualty Care (TCCC) training module; HTC VivePro	Posttest knowledge acquisition	No significant improvement in posttest scores between AR and control groups
Aranda-García et al, 2024 [30]	Health sciences and nursing students (n=60)	CPR and AED[e] training module; Vuzix	Time to task completion, adherence to BLS protocol, CPR performance	Significantly improved CPR quality and protocol adherence in AR group
Follman et al, 2021 [31]	Non-EM[f] health care professionals (n=40)	Real-time assistance in MCI triage; ReconJet	Time to triage; triage accuracy	Significantly decreased triage time in non-AR; no difference in accuracy
Hou et al, 2022 [32]	Health care university students (n=27)	CPR training module; HoloLens	CPR performance metrics (compression rate and depth)	No significant performance difference between AR and control groups
Apiratwarakul et al, 2022 [33]	Emergency physicians, nurses, and EMTs[g] (n=68)	Real-time assistance in MCI casualty detection; HMT-1	Time to completion; accuracy of casualty count in simulated MCI	Significantly decreased time to task completion in AR group, no significant difference in accuracy
Azimi et al, 2018 [34]	EM providers (n=20)	Training in advanced life support procedures; HoloLens	Task performance, task time	No significant difference between AR and control groups
Glick et al, 2021 [35]	Medical students (n=13)	Remote guidance in performing chest thoracotomy; HoloLens	Procedure quality rated by independent observer	Significantly improved procedure quality rating in AR group

[a]CPR: cardiopulmonary resuscitation.

[b]EMS: emergency medical service.

[c]BLS: basic life support.

[d]MCI: mass casualty incident.

[e]AED: automated external defibrillator.

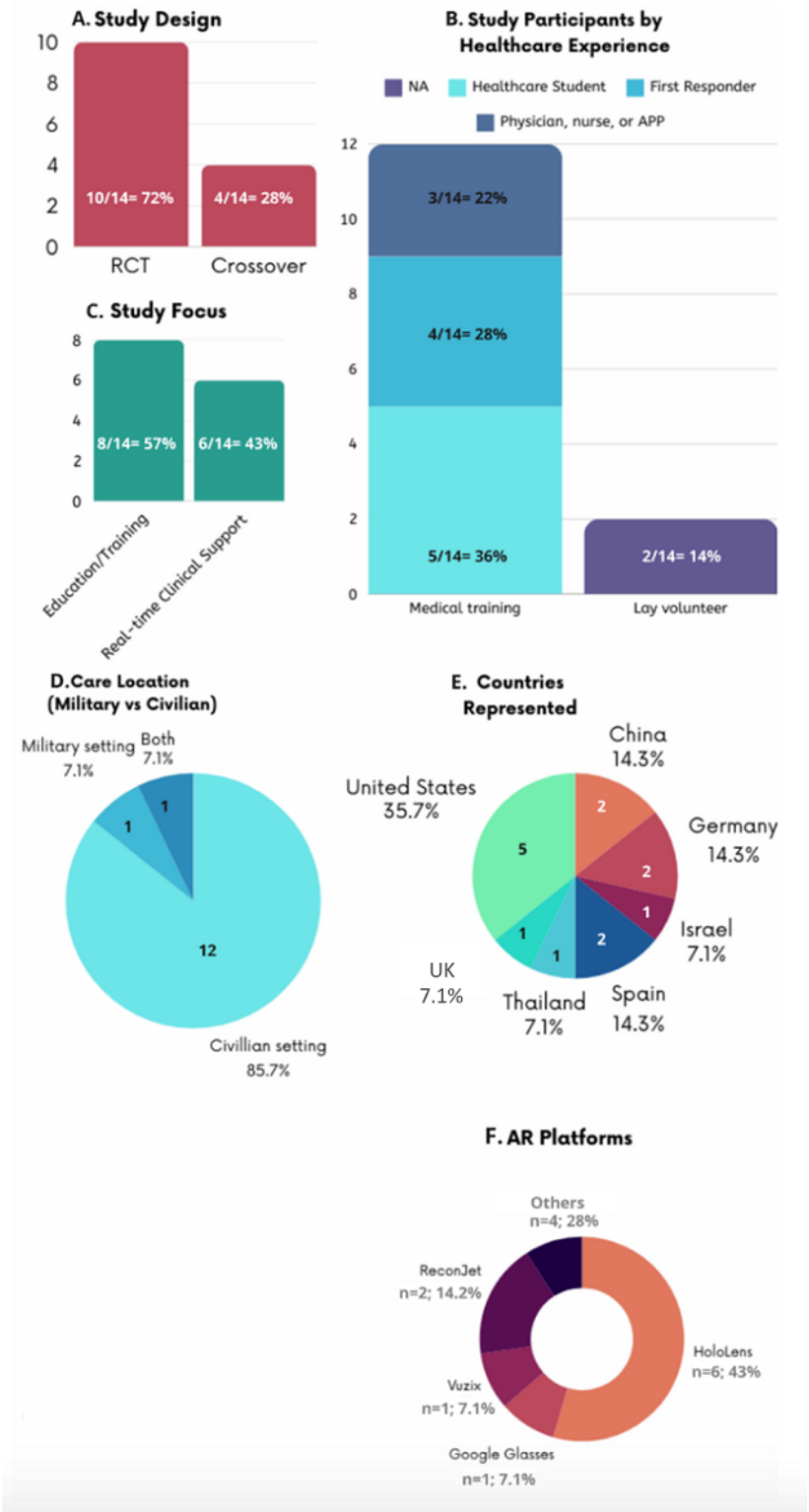
[f]EM: emergency medicine.

[g]EMT: emergency medical technician.

Type of Study Design

Figure 2 highlighted the summary-level study characteristics of the 14 studies. Figure 2A and C shows the distribution of studies by study design (Crossover RCT and Full RCT) and their focus areas: real-time decision support, training or education, or both. Full RCTs are the most frequent, with 4 studies focused on training or education and 3 on real-time decision support. Additionally, one study addressed both focus areas. Crossover RCTs primarily focus on training or education (4 studies), with one study focused on real-time decision support.

Figure 2. Summary characteristics of 14 included studies.



Settings and Regions

The 14 studies included a total of 420 participants and were conducted in 7 different countries. A total of 10 (71%) studies were full RCTs while 4 (29%) studies used a crossover design. Overall, 12 (86%) studies were conducted in civilian settings while 1 (7%) study was conducted in a military setting and 1 (7%) study used both military and civilian settings (Figure 2E). Eight (57%) studies used AR for use in task training and education, while the remaining 6 (43%) used AR to provide real-time decision support for clinical scenarios. All 14 (100%) studies used medical simulation rather than real clinical encounters to test their AR interventions.

Measured Outcomes

While specific outcome measures varied, all studies aimed to compare the efficacy of their AR intervention relative to the current standard of practice. Outcomes examined included time to initiation or completion of desired procedure or intervention (n=5) percentage of correctly informed procedures, procedure quality, or error rate (n=8), and knowledge acquisition (n=1). Overall; 57% (n=8) found statistically significant improvements in their desired outcomes using AR modalities, while 36% (n=5) indicated no significant difference, and 7% (n=1) demonstrated worse performance following AR interventions.

Type of AR Platforms

All studies used wearable head-mounted displays to deliver their AR intervention (Figure 2F). The most used AR platform across studies was HoloLens goggles (6/14; 43%); other AR platforms used included Vuzix (n=1), Google glasses (n=1), ReconJet (n=2), Epson Moverio (n=1), HTC Vive Pro (n=1), and HMT-1 (n=1). A description of the AR platforms used in the 14 studies is presented in Table 2.

A variety of apps and software platforms were used across the 14 studies; selected novel interventions are highlighted in Table 3.

Table . Comparison of augmented reality/virtual reality (AR/VR) devices used in prehospital simulations, showing manufacturer, model, release date, price, and key features.

Device	Manufacturer; models and release date; and retail price	Capabilities
HoloLens	Microsoft; V2 (2019); US $3500	Eye-tracking, audio and speech command, spatial mapping, MR[a] capture, Windows connectivity
Google Glasses	Google X; Explorer (2019) NOTE: no longer manufactured; US $999-US $1848	Voice command, internet browsing, camera, calendar, android iOS
Moverio	Epson; BT 35-e (2018); US $200-US $800	Voice recognition, high definition (HD) display, drone connectivity, remote service, and support
Vuzix	Vuzix; M400 (2020); US $1799	Voice recognition, eye-tracking, spatial mapping, iOS and Android compatibility, waterproof
RealWear	RealWear; HMT-1 (2018); US $797-US $1500	Voice-activated display, noise cancellation, voice-activated, outdoor-compatible display, water and shock resistant, android and Bluetooth compatible, 20-degree field of view
ReconJet	Intel; Smart Glasses (2015); US $699	3-axis sensor, biometric tracking data (heart rate, sleep, etc), GPS, accelerometer, microphones, android iOS compatible, Bluetooth and wifi connectivity
HTC VivePro	HTC; VivePro 2.0 (2021); US $699-US $1999	5k resolution, submillimeter tracking capabilities, balanced ergonomic, 120-degree horizontal field of view

[a]MR: mixed reality.

Table . Selected augmented reality (AR) apps and software platforms in 14 prehospital included studies.

App	Description	Platform (location)	Manuscript
PRIOR	Android app for technical support in MCI[a] triage	Tech2Go GMBH Mobile System (Hamburg, Germany)	Follman et al, 2019 [28]
AUDIME	Android app for technical support in MCI triage in the disaster setting	Tech2Go GMBH Mobile System (Hamburg, Germany)	Follman et al, 2021 [31]
AMBUS	App for learning layout of Ambulance Bus Systems	Unity Game Systems (San Francisco, CA)	Koutitas et al, 2019 [23]
Tensor Flow	Artificial intelligence android app for assistance with casualty detection	Google (Mountain View, CA)	Apiratwarakul et al, 2022 [33]
Juxtopia CAMMRAD PREPARE	App for training in BLS[b] procedures	Juxtopia AR systems (Baltimore, MD)	Collington, 2018 [26]

[a]MCI: mass casualty incident.

[b]BLS: basic life support.

Applications

AR as CDS Tools

A total of 6 studies examined AR-based real-time decision support in the prehospital setting. Rebol et al [22] investigated AR-based real-time feedback for adult cardiopulmonary resuscitation (CPR). They found no significant difference in CPR quality in non–health care university students receiving real-time mixed reality–based feedback on performance as compared with students receiving feedback via standard video conference. Barcala-Furelos et al [27] investigated an AR-based intervention aimed at guiding lifeguards assisting in imminent childbirth situations. They found significantly higher adherence to out-of-hospital birth protocols in the AR-intervention group than in the control group (P<.05 for all protocol variables). Follmann et al [28] found that real-time AR-based guidance in mass casualty incident (MCI) triage led to a significant improvement in triage accuracy over the control group, which performed triage without AR assistance (P=.04). A similar result was found by Follman et al [31], which examined the effect of AR support on MCI triage time and accuracy; they found that triage time was significantly reduced in the control group (P<.001) but found no difference in triage accuracy between groups. Apiratwarakul et al [33] employed an AR intervention for assistance in casualty identification; results demonstrated a decreased time to completion of casualty count in the AR group (P<.05) but no significant difference in accuracy. Glick et al [35] investigated real-time AR-based guidance for medical students in performing a chest thoracotomy and found that expert rating of procedure quality was significantly improved in the AR group (P=.004).

AR as Training Tools

A total of 7 studies examined the utility of AR for education and training in the prehospital setting. Two studies (Doswell et al [25] and Collington et al [26]) investigated AR-augmented training for BLS procedures such as Narcan administration and tourniquet application. Doswell et al [25] found no significant difference in procedure time and accuracy between the AR training group and control group; Collington et al [26] showed an increase in self-reported skills proficiency in the AR training group (mean 2.2, SD 1.03) but no significant difference in clinical proficiency. One study [34] examined the efficacy of an AR-based training module on performing advanced life support procedures, including needle chest decompression, direct intravenous placement, and cricothyroidotomy, but found no significant difference in procedure performance between the AR and standard training groups. One study [23] demonstrated that an AR-based training module for familiarization with an AmBus system led to a 10% reduction in time to task completion (involving finding objects on the AmBus) and 34% reduction in errors than the group receiving standard audiovisual-based training. Two studies (Du et al [29] and Follman et al [31]) examined AR-based training for tactical combat casualty care (TCCC) and MCI triage. Du et al [29], which examined TCCC knowledge gain based on pre and posttraining tests, found no significant performance difference between the AR-based training group and the control group.

A total of 3 studies (Gruenerbl et al [24], Aranda-García et al [30], and Hou et al [32]) specifically examined the performance of adult CPR following AR-based training modules. Two of the 3 studies (Gruenerbl et al [24]; Aranda-García et al [30]) found significant improvement in aspects of CPR performance following AR intervention. They demonstrated a significantly improved percentage of time spent performing chest compressions at the correct depth and rate among nursing students receiving AR-based instruction as compared with standard teaching (P<.001, F=14.85). Aranda-García et al [30] demonstrated significant improvement in the percentage of chest compressions performed with adequate chest recoil (P=.008) among health sciences and nursing students receiving AR-based instruction as compared with control; however, they did not find a significant difference in other metrics. Hou found no significant difference in CPR performance (chest compression rate and depth) receiving AR-based training as compared with instructor-led training.

Risk of Bias Analysis

Risk of bias of studies was assessed via Cochrane's risk of bias tool, which examined parameters including sampling technique, adequacy of randomization, reliability of outcome measures, and statistical power (Multimedia Appendix 3). Overall, the

quality of the included studies was judged to be high. Each of the 14 studies was examined on a manuscript level with consensus reached between 8 independent reviewers. All 14 studies were determined to have a randomized design, with 10 comprising full RCTs and 4 having a crossover design. Most studies were found to have adequate randomization methodology, similar baseline participant characteristics, reliable outcome measures, and a participant dropout rate below 20%. Two of the 14 studies were recorded as lacking sufficient sample size to achieve 80% power with one recorded as "unable to be determined."

Discussion

Principal Findings

This systematic review sought to examine the application of AR to emergency medical care in the prehospital setting, with the primary objective of evaluating the efficacy or effectiveness of AR apps in improving patient outcomes, care processes, and learning outcomes. Of the 14 studies analyzed in this systematic review, the majority demonstrated a significant improvement in desired outcomes with the integration of AR into their workflow, suggesting that AR may have a valuable role to play in enhancing the quality of prehospital care.

AR as CDS Tools

Studies investigating the utility of AR in providing real-time CDS demonstrated a significant improvement in at least 1 outcome. AR interventions are especially effective in providing real-time decision support for MCI scenarios, enhancing both the accuracy and efficiency of triage procedures and casualty counts. AR-based remote guidance improved procedure quality for fully-trained medical students performing simulated chest thoracotomy procedures, as well as for laypeople responding to simulated childbirth. These results suggest that AR may have an important role to play in improving medical control for EMS, as AR-based feedback and guidance could greatly enhance decision-making for prehospital care providers as compared with traditional audio feedback [36-38]. Results of these studies also suggest that AR may serve a vital purpose in tactical emergency medicine scenarios, including military and law enforcement operations that could benefit from remote guidance in high-acuity scenarios [35,39]. Future research could investigate AR integration into tactical emergency medicine scenarios, such as SWAT team activations.

It is also important to note the potential integration of AR with other emerging technologies, such as artificial intelligence algorithms, which could further enhance decision support by providing predictive analytics and personalized recommendations [13,40,41]. Combining AR with wearable biometric sensors could offer real-time monitoring of vital signs, providing a context-aware decision support system that enhances situational awareness and operational efficiency [10].

AR as Training Tools

With regards to education and training, 2 of the 4 studies examining the benefit of AR in augmenting CPR training demonstrated significant improvement in CPR quality following AR intervention. These findings suggest that it may be feasible to integrate AR into CPR training. The study by Koutitas et al [23], which examined an AR-based training module for familiarization with AmBus systems also demonstrated improved task completion and enhanced comfort and familiarity with the vehicle in the AR intervention group, suggesting that AR may prove a useful adjunct to EMS companies in training new hires. Notably, some studies, that examined AR intervention in prehospital education and training modules for skills including, CPR, BLS, advanced life support procedures, and TCCC, showed no difference in performance with AR intervention. It is possible that some of these tasks, which involve a significant number of hands-on skills, were more difficult to adapt from in-person instruction to AR-based training. Future research could more thoroughly explore discrepancies in AR-based training modules among various prehospital clinical skills [42]. Furthermore, the scalability of AR training modules offers a significant advantage for widespread training initiatives, allowing consistent and repeatable training experiences across different geographical locations. This scalability is particularly beneficial for remote and underserved areas where access to high-quality training resources is limited.

Challenges of AR Technology

Overall satisfaction with AR platforms was high across the 14 studies; manuscripts that solicited user feedback found that most participants reported positive perceptions of the technology. Several common concerns emerged from this user feedback. These common concerns are summarized in Table 4.

Of greatest concern was user comfort as well as occasional unpleasant side effects associated with the use of AR. Several manuscripts indicated that wearable interventions, particularly those including headsets, were not compatible with participants who wore prescription eyeglasses. Additionally, some reported participants experiencing side effects after AR use, including dizziness, headache, and nausea. This constellation of adverse effects is collectively known as "cybersickness [43]," and has been demonstrated to impact AR, mixed reality, and virtual reality users, particularly those who are susceptible to motion sickness [44]. Future research into AR should factor cybersickness risk into study design and look to mitigate side effects. Other common concerns included the costs associated with both the purchase and maintenance of AR platforms [45], as well as inconsistent user interface and frequent technological glitches [46]. Addressing these concerns requires a multi-faceted approach [47,48]. Collaborations with manufacturers, health care providers, and end users will be crucial in creating AR systems that are not only effective but also user-friendly and economically viable [9,12]. Additionally, ongoing education and support for users can help mitigate some of the initial discomfort and resistance to new technology [49].

Table . Summary of common concerns related to augmented reality (AR) use in prehospital care, including user comfort, user interface issues, information technology (IT) challenges, and cost.

Concern	Source
User comfort	• Headgear uncomfortable or disruptive to workflow, causes unpleasant side effects (Rebol et al, 2023 [22]; Doswell et al, 2020 [25]; Follman et al, 2019 [28]; Du et al, 2022 [29]; Follman et al, 2021 [31]; Hou et al, 2022 [32]) • AR implicated: HoloLens, Google Glass, Moverio
User interface	• User interface confusing or difficult to use or requires steep learning curve (Follman et al, 2021 [31]; Glick et al, 2021 [35]) • AR implicated: HoloLens, ReconJet
IT issues	• Poor battery life, screen glitching, application freezing (Rebol et al, 2023 [22]; Barcala-Furelos et al, 2023 [27]; Aranda-García et al, 2024 [30]; Follman et al, 2021 [31]) • AR implicated: HoloLens, ReconJet, Vuzix
Cost	• High cost of materials, setup, and maintenance (Du et al, 2022 [29]) • AR implicated: HTC VivePro

Limitations and Future Directions

This systematic review had several limitations. First, many of the included studies were of small sample size. Most studies included under 50 participants, with several included 10 or fewer, which may result in some included studies being underpowered. It is not unusual for studies investigating expensive technologies in potentially cumbersome settings to by necessity include small numbers; however, future research can prioritize adequate sample sizes to ensure robust statistical analyses. Second, our review compared studies with variable outcomes and statistical methodology and thus was not able to examine data in aggregate. A potential next step would be to conduct a meta-analysis of AR interventions in specific emergency prehospital applications, such as CPR training or MCI triage. Third, this review only included studies of AR apps in the prehospital care of adults. Future research will include inquiries into applications of AR for use with pediatric populations. Finally, a main limitation of our search approach was the potential for missed manuscripts due to not features like MeSH headers in PubMed. However, the use of broad search terms across multiple databases helped mitigate this limitation.

Conclusion

This systematic review shows the promising role of AR technology in enhancing the efficacy of prehospital emergency care. The analyzed studies, involving a total of 14 RCTs demonstrate that AR may enhance clinical decision-making and training modalities within prehospital settings. These improvements are crucial in high-stakes environments where rapid and accurate response is essential. Challenges related to technology integration, cost, and user acceptance remain. Addressing these barriers and conducting further research will be vital for realizing the full potential of AR in prehospital care delivery.

Acknowledgments

This research received funding through the Stepping Strong Innovator Awards, an initiative by the Stepping Strong Center for Trauma Innovation, Harvard Medical School, at Mass General Brigham. This research did not use generative artificial intelligence tools for data analysis. Generative artificial intelligence (specifically ChatGPT) was used for revising the abstract.

Data Availability

This study is a systematic review, and all data analyzed are derived from previously published studies. Complete references and sources for all data used are provided within the paper, ensuring full transparency and accessibility.

Conflicts of Interest

AE serves as an advisor for MedVR Education and Apoqlar. Their products are not discussed in this paper.

Multimedia Appendix 1
Detailed search strategy across databases for identifying studies on augmented reality in prehospital emergency care.
[DOCX File, 21 KB - xr_v2i1e66222_app1.docx]

Multimedia Appendix 2
Systematic review form used for extraction relevant information from included papers.
[DOCX File, 23 KB - xr_v2i1e66222_app2.docx]

Multimedia Appendix 3
Bias evaluation tool questions.
[DOCX File, 17 KB - xr_v2i1e66222_app3.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.
[DOCX File, 32 KB - xr_v2i1e66222_app4.docx]

References

1. Vukmir RB. Survival from prehospital cardiac arrest is critically dependent upon response time. Resuscitation 2006 May;69(2):229-234. [doi: 10.1016/j.resuscitation.2005.08.014] [Medline: 16500015]
2. Thind A, Hsia R, Mabweijano J, Hicks ER, Zakariah A, Mock CN. Prehospital and emergency care. In: Debas HT, Donkor P, Gawande A, Jamison DT, Kruk ME, Mock CN, editors. Essential Surgery: Disease Control Priorities, 3rd edition: The International Bank for Reconstruction and Development / The World Bank; 2015:245-262. [doi: 10.1596/978-1-4648-0346-8_ch14]
3. Blackwell TH. Prehospital care. Emerg Med Clin North Am 1993 Feb;11(1):1-14. [doi: 10.1016/S0733-8627(20)30653-2] [Medline: 8432242]
4. Lombardi G, Gallagher J, Gennis P. Outcome of out-of-hospital cardiac arrest in New York City. The Pre-Hospital Arrest Survival Evaluation (PHASE) Study. JAMA 1994 Mar 2;271(9):678-683. [Medline: 8309030]
5. Regel G, Stalp M, Lehmann U, Seekamp A. Prehospital care, importance of early intervention on outcome. Acta Anaesthesiol Scand 1997 Jun;41(S110):71-76. [doi: 10.1111/j.1399-6576.1997.tb05508.x]
6. Sayre MR, White LJ, Brown LH, McHenry SD, National EMS Research Strategic Plan Writing Team. The national EMS research strategic plan. Prehosp Emerg Care 2005;9(3):255-266. [doi: 10.1080/10903120590962238] [Medline: 16147473]
7. Snooks H, Evans A, Wells B, et al. What are the highest priorities for research in emergency prehospital care? Emerg Med J 2009 Aug;26(8):549-550. [doi: 10.1136/emj.2008.065862] [Medline: 19625545]
8. Holroyd BR, Knopp R, Kallsen G. Medical control. Quality assurance in prehospital care. JAMA 1986;256(8):1027-1031. [doi: 10.1001/jama.256.8.1027] [Medline: 3525878]
9. Beadle S, Spain R, Goldberg B, et al. Virtual reality, augmented reality, and virtual environments: demonstrations of current technologies and future directions. Proc Hum Factors Ergon Soc Annu Meet 2020 Dec;64(1):2119-2123. [doi: 10.1177/1071181320641514]
10. Ebnali M, Goldsmith A, Burian B, et al. AR-coach: using augmented reality (AR) for real-time clinical guidance during medical emergencies on deep space exploration missions. Presented at: 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022); Jul 24-28, 2022; New York City, NY, United States. [doi: 10.54941/ahfe1002100]
11. Ebnali M, Paladugu P, Miccile C, et al. Extended reality applications for space health. Aerosp Med Hum Perform 2023 Mar 1;94(3):122-130. [doi: 10.3357/AMHP.6131.2023] [Medline: 36829279]
12. Burian BK, Ebnali M, Robertson JM, et al. Using extended reality (XR) for medical training and real-time clinical support during deep space missions. Appl Ergon 2023 Jan;106:103902. [doi: 10.1016/j.apergo.2022.103902] [Medline: 36162274]
13. Harari R, Al-Taweel A, Ahram T, Shokoohi H. Explainable AI and augmented reality in transesophageal echocardiography (TEE) imaging. Presented at: 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR); Jan 17-19, 2024; Los Angeles, CA, United States. [doi: 10.1109/AIxVR59861.2024.00051]
14. Seetohul J, Shafiee M, Sirlantzis K. Augmented reality (AR) for surgical robotic and autonomous systems: state of the art, challenges, and solutions. Sensors (Basel) 2023 Jul 6;23(13):6202. [doi: 10.3390/s23136202] [Medline: 37448050]
15. Gholizadeh M, Bakhshali MA, Mazlooman SR, et al. Minimally invasive and invasive liver surgery based on augmented reality training: a review of the literature. J Robot Surg 2023 Jun;17(3):753-763. [doi: 10.1007/s11701-022-01499-2] [Medline: 36441418]
16. Sugimoto M, Yasuda H, Koda K, et al. Image overlay navigation by markerless surface registration in gastrointestinal, hepatobiliary and pancreatic surgery. J Hepatobiliary Pancreat Sci 2010 Sep;17(5):629-636. [doi: 10.1007/s00534-009-0199-y] [Medline: 19798463]
17. Vávra P, Roman J, Zonča P, et al. Recent development of augmented reality in surgery: a review. J Healthc Eng 2017;2017:4574172. [doi: 10.1155/2017/4574172] [Medline: 29065604]
18. Elsayed M, Kadom N, Ghobadi C, et al. Virtual and augmented reality: potential applications in radiology. Acta Radiol 2020 Sep;61(9):1258-1265. [doi: 10.1177/0284185119897362] [Medline: 31928346]

19. Uppot RN, Laguna B, McCarthy CJ, et al. Implementing virtual and augmented reality tools for radiology education and training, communication, and clinical care. Radiology 2019 Jun;291(3):570-580. [doi: 10.1148/radiol.2019182210] [Medline: 30990383]
20. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Rebol M, Steinmaurer A, Gamillscheg F, et al. CPR emergency assistance through mixed reality communication. arXiv. Preprint posted online on Dec 14, 2023. [doi: 10.1007/978-3-031-32883-1_38]
23. Koutitas G, Smith KS, Lawrence G, et al. A virtual and augmented reality platform for the training of first responders of the ambulance bus. Presented at: PETRA '19; Jun 5-7, 2019; Rhodes, Greece. [doi: 10.1145/3316782.3321542]
24. Gruenerbl A, Javaheri H, Monger E, Gobbi M, Lukowicz P. Training CPR with a wearable real time feedback system. Presented at: UbiComp '18; Oct 8-12, 2018; Singapore. [doi: 10.1145/3267242.3267277]
25. Doswell JT, Johnson J, Brockington B, Mosby A, Chinery A. Juxtopia ® CAMMRAD PREPARE: wearable AI-AR platform for clinical training emergency first response teams. Presented at: 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR); Dec 14-18, 2020; Utrecht, Netherlands. [doi: 10.1109/AIVR50618.2020.00047]
26. Collington D, Doswell J, Johnson J, Messam M, Salaam S, Boykin A. The pervasive rapid E-preparedness augmented reality environment (PREPARE) E-learning intervention for HAZWOPER training of fire fighter/emts. Presented at: 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT); Jul 9-13, 2018; Mumbai, India. [doi: 10.1109/ICALT.2018.00096]
27. Barcala-Furelos R, Aranda-García S, Otero-Agra M, et al. Are smart glasses feasible for dispatch prehospital assistance during on-boat cardiac arrest? A pilot simulation study with fishermen. Intern Emerg Med 2023 Aug;18(5):1551-1559. [doi: 10.1007/s11739-023-03251-6] [Medline: 37014496]
28. Follmann A, Ohligs M, Hochhausen N, Beckers SK, Rossaint R, Czaplik M. Technical support by smart glasses during a mass casualty incident: a randomized controlled simulation trial on technically assisted triage and telemedical app use in disaster medicine. J Med Internet Res 2019 Jan 3;21(1):e11939. [doi: 10.2196/11939] [Medline: 30609988]
29. Du W, Zhong X, Jia Y, et al. A novel scenario-based, mixed-reality platform for training nontechnical skills of battlefield first aid: prospective interventional study. JMIR Serious Games 2022 Dec 6;10(4):e40727. [doi: 10.2196/40727] [Medline: 36472903]
30. Aranda-García S, Otero-Agra M, Berlanga-Macías C, et al. New communication tool for basic life support training: smart glasses. A quasi-experimental study. Medicina Intensiva (Engl Ed) 2024 Feb;48(2):77-84. [doi: 10.1016/j.medine.2023.10.011]
31. Follmann A, Ruhl A, Gösch M, Felzen M, Rossaint R, Czaplik M. Augmented reality for guideline presentation in medicine: randomized crossover simulation trial for technically assisted decision-making. JMIR Mhealth Uhealth 2021 Oct 18;9(10):e17472. [doi: 10.2196/17472] [Medline: 34661548]
32. Hou L, Dong X, Li K, et al. Comparison of augmented reality-assisted and instructor-assisted cardiopulmonary resuscitation: a simulated randomized controlled pilot trial. Clin Simul Nurs 2022 Jul;68:9-18. [doi: 10.1016/j.ecns.2022.04.004] [Medline: 35646196]
33. Apiratwarakul K, Cheung LW, Tiamkao S, et al. Smart glasses: a new tool for assessing the number of patients in mass-casualty incidents. Prehosp Disaster Med 2022 Aug;37(4):480-484. [doi: 10.1017/S1049023X22000929] [Medline: 35757837]
34. Azimi E, Winkler A, Tucker E, et al. Can mixed-reality improve the training of medical procedures? Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:4065-4068. [doi: 10.1109/EMBC.2018.8513387] [Medline: 30441249]
35. Glick Y, Avital B, Oppenheimer J, et al. Augmenting prehospital care. BMJ Mil Health 2021 Jun;167(3):158-162. [doi: 10.1136/jramc-2019-001320] [Medline: 32086268]
36. Demir F. A next-generation augmented reality platform for mass casualty incidents (MCI). J Usability Stud 2017;12(4). [doi: 10.5555/3190867.3190871]
37. Paravantes G, Leehane H, Premarathna D, Chung C, Kappen DL. Application of augmented reality for multi-scale interactions in emergency vehicles. Presented at: AutomotiveUI '18; Sep 23-25, 2018; Toronto, ON, Canada. [doi: 10.1145/3239092.3267415]
38. Koutitas G, Smith S, Lawrence G. Performance evaluation of AR/VR training technologies for EMS first responders. Virtual Real 2021 Mar;25(1):83-94. [doi: 10.1007/s10055-020-00436-8]
39. Munzer BW, Khan MM, Shipman B, Mahajan P. Augmented reality in emergency medicine: a scoping review. J Med Internet Res 2019 Apr 17;21(4):e12368. [doi: 10.2196/12368] [Medline: 30994463]
40. Ebnali M, Ahmadi N, Nabiyouni E, Karimi H. AI-powered human digital twins in virtual therapeutic sessions. Proc Int Symp Hum Factors Ergon Health Care 2023 Mar;12(1):1-4. [doi: 10.1177/2327857923121000]
41. Battineni G, Chintalapudi N, Ricci G, Ruocco C, Amenta F. Exploring the integration of artificial intelligence (AI) and augmented reality (AR) in maritime medicine: a narrative literature review. JMIR Preprints. Preprint posted online on Oct 22, 2023. [doi: 10.2196/preprints.53870]

42. Zhu E, Hadadgar A, Masiello I, Zary N. Augmented reality in healthcare education: an integrative review. PeerJ 2014;2:e469. [doi: 10.7717/peerj.469] [Medline: 25071992]
43. Jeon H. Factors affecting motion sickness in an augmented reality environment. Presented at: The 13th International Conference on Advances in Computer-Human Interactions; Nov 21-25, 2020; Valencia, Spain.
44. Spilka D. Is virtual reality bad for our health? studies point to physical and mental impacts of VR usage. Springer Nature. 2023. URL: https://socialsciences.nature.com/posts/is-virtual-reality-bad-for-our-health-studies-point-to-physical-and-mental-impacts-of-vr-usage [accessed 2025-01-20]
45. Ara J, Karim FB, Alsubaie MSA, et al. Comprehensive analysis of augmented reality technology in modern healthcare system. IJACSA 2021;12(6). [doi: 10.14569/IJACSA.2021.0120698]
46. Oke AE, Arowoiya VA. Critical barriers to augmented reality technology adoption in developing countries: a case study of Nigeria. JEDT 2022 Aug 30;20(5):1320-1333. [doi: 10.1108/JEDT-12-2020-0519]
47. Eder PA, Rashid A. Telemedicine for prehospital trauma care: a promising approach. In: The High-Risk Surgical Patient: Springer International Publishing; 2023:683-689. [doi: 10.1007/978-3-031-17273-1_61]
48. Borna A, Mousavi SZ, Fathollahzadeh F, Nazeri A, Harari RE. Applications of augmented and virtual reality in enhancing communication for individuals who are hard of hearing: a systematic review. Am J Audiol 2024 Dec 2;33(4):1378-1394. [doi: 10.1044/2024_AJA-24-00056] [Medline: 39374491]
49. Schlosser P. Head-worn displays for emergency medical services. Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411763.3443421]

Abbreviations

ACM: Association for Computing Machinery
AR: augmented reality
BLS: basic life support
CDS: clinical decision support
CPR: cardiopulmonary resuscitation
EM: emergency medicine
EMT: emergency medical technician
IEEE: Institute of Electrical and Electronics Engineers
MCI: mass casualty incident
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
TCCC: tactical combat casualty care

Edited by A Coristine; submitted 07.09.24; peer-reviewed by M Rubin; revised version received 27.11.24; accepted 27.11.24; published 11.02.25.

Please cite as:
Harari RE, Schulwolf SL, Borges P, Salmani H, Hosseini F, Bailey SKT, Quach B, Nohelty E, Park S, Verma Y, Goralnick E, Goldberg SA, Shokoohi H, Dias RD, Eyre A
Applications of Augmented Reality for Prehospital Emergency Care: Systematic Review of Randomized Controlled Trials
JMIR XR Spatial Comput 2025;2:e66222
URL: https://xr.jmir.org/2025/1/e66222
doi:10.2196/66222

Understanding the Views of Health Care Professionals on the Usability and Utility of Virtual Reality Multidisciplinary Team Meetings: Usability and Utility Study

Maryam Almashmoum[1,2], BSc, MSc; Antony Payton[1,3], MSc, PhD; Emily Johnstone[1,3], MSc, BSc; James Cunningham[1], MSc, PhD; John Ainsworth[1,4], MSc, PhD

[1]Division of Informatics Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine, and Health, University of Manchester, Vaughan House, Portsmouth St., Manchester, United Kingdom

[2]Nuclear Medicine Department, Faisal Sultan Bin Eissa, Kuwait Cancer Control Center, Kuwait City, Kuwait

[3]VREvo Ltd, The University of Manchester Core Technology Facility, Manchester, United Kingdom

[4]NIHR Manchester Biomedical Research Centre, Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom

Corresponding Author:
Maryam Almashmoum, BSc, MSc
Division of Informatics Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine, and Health, University of Manchester, Vaughan House, Portsmouth St., Manchester, United Kingdom

Abstract

Background: Multidisciplinary team (MDT) meetings are one of the facilitators that enhance knowledge sharing among health care professionals. However, organizing a face-to-face MDT meeting to discuss patient treatment plans can be time-consuming. Virtual reality software is widely used in health care nowadays to save time and protect lives. Therefore, the use of virtual reality multidisciplinary team (VRMDT) meeting software may help enhance knowledge sharing between health care professionals and make meetings more efficient.

Objective: The objectives of this study were to introduce VRMDT software for enhancing knowledge sharing and to evaluate the feasibility and usability of the VRMDT for use by professionals in health care institutions.

Methods: We invited participants from The University of Manchester Faculty for Biology, Medicine, and Health who had a health care background. As this was the first stage of software development, individuals who did not usually attend MDT meetings were also invited via email to participate in this study. Participants evaluated VRMDT using a Meta Quest 3 headset, and software developed using the Unity platform. The software contained an onboarding tutorial that taught the participants how to select items, load and rotate 3D Digital Imaging and Communications in Medicine files, talk to a generative artificial intelligence–supported avatar, and make notes. After the evaluation (approximately 15 min), participants received an electronic survey using the Qualtrics survey tool (Qualtrics International Inc) to score the usability and feasibility of the software by responding to the 10-item system usability scale, and 12-point heuristic evaluation questions with Neilsen severity rating.

Results: A total of 12 participants, including 4 health informatics, 3 with a nursing background, 2 medical doctors, 1 radiologist, and 2 biostatisticians, participated in the study. The most common age bracket of participants was 20 - 30 years (6/12, 50%). Most of the respondents had no experience with virtual reality, either in educational or entertainment settings. The VRMDT received a mean usability score of 72.7 (range between 68 and 80.3), earning an overall “good” rating grade. The mean score of single items in the heuristic evaluation questionnaires was less than 1 out of 4 (the overall mean was 0.6), which indicates that only minor problems were encountered when using this software. Overall, the participant’s feedback was good with highlighted issues including a poor internet connection and the quality of the generative artificial intelligence response.

Conclusions: VRMDT software (developed by Sentira^{XR}) was developed with several functions aimed at helping health care professionals to discuss medical conditions efficiently. Participants found that the VRMDT is a powerful, and useful tool for enhancing knowledge sharing among professionals who are involved in MDT meetings due to its functionality and multiuser interactive environments. Additionally, there may be the possibility of using it to train junior professionals to interpret medical reports.

(JMIR XR Spatial Comput 2025;2:e60651) doi:10.2196/60651

KEYWORDS

knowledge sharing; multidisciplinary team meetings; artificial intelligence; heuristic evaluation; usability; virtual reality; VR; simulation; virtual environments; digital environments

Introduction

Overview

The United Kingdom's health care sector is facing significant pressures from increased patient demands and workforce supply issues. A need for efficiently connected health care employees is important for sharing knowledge and it is an integral part of knowledge management. During COVID-19, communication across sectors moved towards web-based communication methods [1-3], such as videoconferencing (eg, Microsoft Teams and Zoom), which helped to protect the lives of patients and staff [3-5]. To maintain knowledge-sharing practices among professionals, there are several professional digital communities [6,7]. The purpose of these professional digital communities is to get professionals with common expertise to share their knowledge without considering geographical barriers [6]. Virtual multidisciplinary team (MDT) meetings have been shown to have a visible role in maintaining communication among cancer care professionals to discuss, follow up, and set clear treatment plans [8]. Additionally, it has been shown to improve cancer patients' outcomes [9-13]. Traditional face-to-face methods of MDT have drawbacks that limit attendance including lack of time and funding [8]. Introducing new technology for communication has benefits, although there are also drawbacks such as reliance on bandwidth, increased conversation time, and loss of gesture communication that can be difficult compared with traditional methods, thereby directly affecting good decision-making [14,15].

The use of videoconferencing has surged as a communication method during and post-COVID, although it has limitations including the inability for natural F-2-F interaction due to the participants only seeing a video image. Additionally, smooth and stable internet network is required to ensure that video conferencing runs smoothly. Moreover, the inability to show 3D images compared with the virtual reality (VR) tools may be a distinct disadvantage [16]. As a result, the existence of a powerful web-based tool that simulates a real environment may have benefits. VR and augmented reality are increasingly being used in the medical field both for training and as a procedural aid [17]. VR is defined as "a three-dimensional computer-generated simulated environment, which attempts to replicate real world or imaginary environments and interactions, thereby supporting work, education, recreation, and health" [3,18]. In addition, the user can interact with avatars using generative artificial intelligence (AI) supported natural language processing (NLP) which further enhances the realism of the experience. It requires head-mounted displays, and either hand controllers or hand tracking in order to perform practical procedures [19]. The sense of presence is one of the key characteristics of VR that makes it different from other communication mediums [14]. The use of VR applications in the health care market has grown massively in recent years. In 2022, the VR health care market reached over US $2.3 billion worldwide, with 171 million VR users [20].

VR in health care has several benefits, such as facilitating training, education, and the development of technical skills. Additionally, VR is being used for a variety of purposes, including surgery and treatment, training, and patient therapy and rehabilitation [21]. Kyaw et al [22], illustrated that using VR applications improves professionals' skills, and knowledge compared with face-to-face communication and web-based digital education. In particular, it has the ability to negate the need for face-to-face contact, while maintaining the illusion of being with colleagues in the real world [23].

There are several factors that affect knowledge sharing in the medical imaging department at cancer centers, which are similar to those in most health care sectors [24]. MDTs are considered important departmental facilitators that enhance knowledge sharing among health care professionals [24]. MDT is considered a pillar of the best practices in cancer canters and plays an important role in cancer Treatment [25]. The United Kingdom's National Health Service definition of MDT is "a group of professionals from one or more clinical disciplines who together make decisions regarding the recommended treatment of individual patients" [26]. MDT in cancer centers is defined as the collaboration of several health care professionals in different fields engaged in the treatment of cancer with the overall objective of enhancing the rate of interpreting treatments of cancer patients, and patient care [13,26]. Cancer centers began to use a multidisciplinary approach in the mid-1980s, and by the 1990s, the MDT meeting was introduced as an instrument for providing coordinated, collaborative care, which allow a broader range of opinions on treatment plans [13,27]. In addition, it provides training for junior health care professionals. However, there are several barriers that contribute to not attending those meetings as per policy recommendations. These include time constraints, lack of departmental arrangements, geographical barriers among health care professionals, and shortage of staff [13].

In health care institutions, implementing new interventions such as VR among health care professionals may overcome current barriers and enhance knowledge-sharing practices to increase patients' outcomes and minimize medical mistakes. However, there are several challenges to implementing VR as a communication tool, including providing evidence that these technologies can save time, increase productivity, and reduce carbon footprint, without adding significant hardware costs and training time [28-30]. The aim of this research is to introduce new technology and perform a usability study of VR in MDT to investigate the feasibility and usability of using VR in cancer health care meetings.

Objectives

In this study, we developed a virtual reality multidisciplinary team (VRMDT) for enhancing communication with professionals, which was evaluated in terms of its usability by professionals from a variety of backgrounds.

The aim of this study was to investigate the usability of newly developed VRMDT software that helps gather health professionals in a 3D immersive environment to aid communication and set a clear treatment plan for the cancer patient. The objectives of this study were:

- To introduce VRMDT software to health care professionals.
- Evaluate the usability, feasibility, and efficacy of VRMDT by applying the System Usability Scale (SUS), and identifying the problems with the user interface by using a heuristic evaluation questionnaire.
- Identify the strengths and weaknesses of using VRMDT.
- Determine if this technology has the potential to increase the number of MDT meetings in cancer centers locally and internationally.
- Increase awareness of using VR technology among health care professionals in cancer centers.

Methods

An Overview of VRMDT Software

The software was designed by our University of Manchester research team and developed using the Unity platform by Sentira[XR] [31], which is a University of Manchester spinout that uses VR and generative AI NLP to create authentic training simulations for health care professionals and other disciplines. The designs of the VRMDT comprise:

- An onboarding section for those not familiar with VR.
- Options to select a health care uniform of varying color and add the name to be displayed above the head of each user's avatar.
- 3D VR meeting room with round table.
- Ability to display a 3D Digital Imaging and Communications in Medicine (DICOM) scan image in the middle of the virtual table to allow 3D visualization. Additionally, there is a screen in front of each user to few the DICOM images in a traditional 2D mode.
- A whiteboard for writing notes and drawing images.
- A laser pointer beside each user for pointing to specific locations on the 3D DICOM images.
- An interactive avatar that uses generative AI NLP to provide answers to questions from users in the room related to the patient's scans, condition, and patient history.
- A master control panel where patient DICOM images can be selected.

The VRMDT (Figure 1) is designed to allow health care professionals to treatment plan anywhere and at any time. To run the VRMDT simulation, a reasonable Wi-Fi connection (≥10 Mbps), head-mounted display, and controllers are required.

Before entering the MDT room, the user had the option to undertake an onboarding scenario that introduced them to basic functionality such as picking up objects, talking to the avatar, selecting DICOM files, and making notes on a whiteboard. The user can then begin the simulation first by typing in their username (displayed over the head of their avatar) and selecting their outfit's color (Figure 2). In the VRMDT software, there is a round table fitting 10 users with a control screen that contains the setting options, selecting the patient DICOM files, and the option to move the control panel to another user. Another screen available to all 10 users displays the traditional 2D DICOM images for cancer patients (Figure 3B). Additionally, the meeting room contains a whiteboard to allow the user to make notes or draw diagrams (Figure 4B). In the middle of the meeting table, the 3D DICOM (Figure 4A) images appear with the facilities to rotate the images on the x-axis to help show any tumors or lesions. A laser pointer is available to each participant to help highlight a region on the 3D image (Figure 3A). DICOM images were retrieved from The Cancer Imaging Archive which are accessible for the public to download and use without ethical approval. The time zones for both the United Kingdom and Kuwait are displayed on the wall of the meeting room.

Generative AI NLP used the InWorld platform [32]. Voice cloning (voice of MA cloned) uses Eleven Labs software which is supported by InWorld [32,33]. Patient information and avatar background details were entered into InWorld and quality assurance was conducted to ensure that the responses from the generative AI NLP had an accuracy of 95% or greater. The generative AI NLP-supported avatar was placed in the meeting room (Figure 5) and allowed the user to ask questions regarding the medical condition of the patients. The Photon platform was used to allow users to speak with each other as they would with any teleconference software [34]. The purpose of the AI-supported avatar was to provide the MDT with specific details on each of the patients, such as name, age, status of the medical condition, medications, chemotherapy/radiotherapy received, response to treatments/medications, bloodwork, and patient concerns. Providing patient information via an avatar, removed the need for reading extensive text notes which is not ideal in a VR environment due to reduced visual resolution and an increased risk of cybersickness. It also allowed for one or more of the MDT to be absent and still provide the information.

For the implantation, the software required a direct connection with the Picture and Archiving and Communication System to visualize patient images. Additionally, the VRMDT contains instructions voiced over to guide the user throughout testing the software.

Figure 1. The environment of the virtual reality multidisciplinary team software.

Figure 2. "On boarding" interface page for selection of the outfits, and the info that will appear on the user (such as name).

Figure 3. (**A**) Two screens: a controlled screen and a screen to display the traditional 2D scan images. (**B**) Laser pointer.

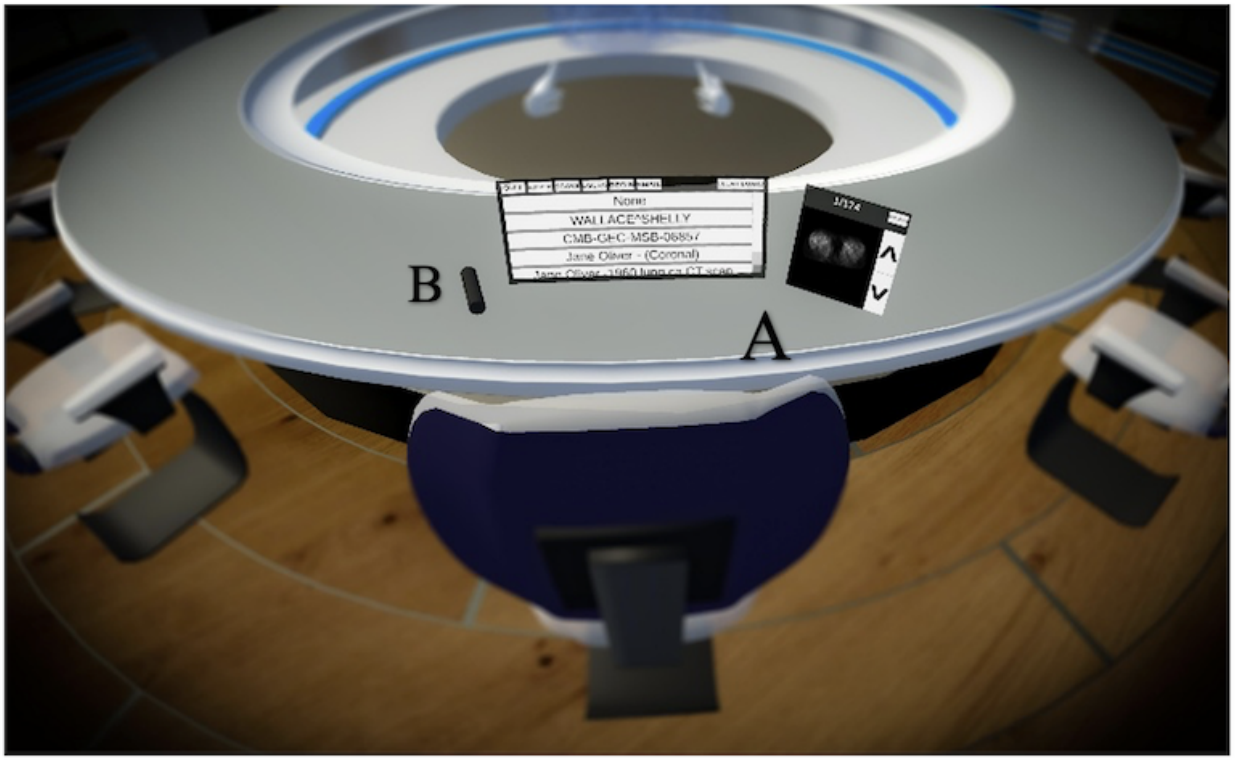


Figure 4. (**A**) The 3D Digital Imaging and Communications in Medicine (DICOM) images and (**B**) a whiteboard.

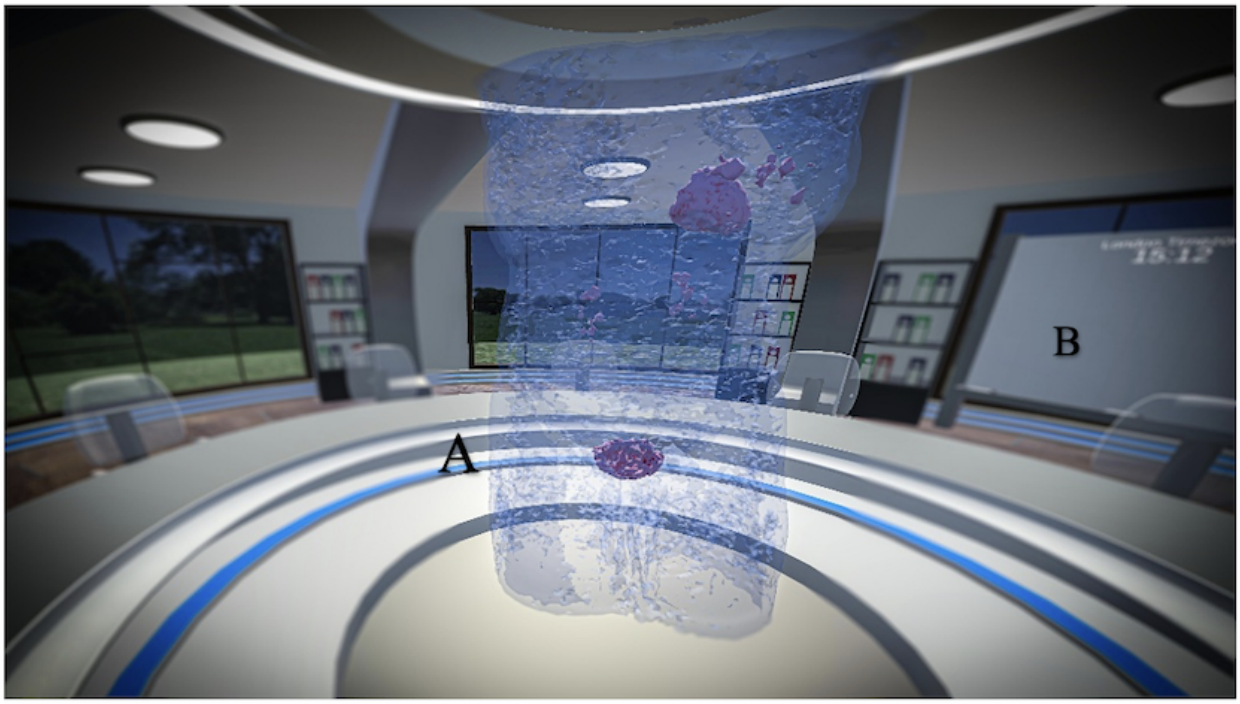


Figure 5. Interactive avatar.

Participants

To be eligible for participation in this study, the participant had to have a health care background, with those recruited being postgraduate students and staff at The University of Manchester. As this was the first stage of software development, participants who were not routinely involved in MDTs were also invited to evaluate the software.

Participants were recruited via email with the inclusion criteria as provided in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Postgraduate students and staff at the University of Manchester.
- 21 years or older.
- Any gender.
- Health care professional background (including but not limited to doctors, nurses, and radiologists).
- Health care professionals who are involved in multidisciplinary teams.
- Willing to provide informed consent.
- English speakers.
- No pre-existing conditions that may cause discomfort or distress in a virtual reality (VR) environment.

Exclusion criteria

- People who do not read, speak or understand English, because the software is in English only.
- People who are unwilling to wear a VR headset.
- People who had a pre-existing condition that may cause discomfort or distress in a VR environment

Instruments

Validated usability and utility questionnaires were used to assess the simulation's efficacy, efficiency, and user pleasure [35]. Two methods were used to assess the usability evaluation: 10-item SUS, and 12-item heuristic evaluation questionnaires [36,37]. Upon completion of the trial, the SUS and heuristic questionnaire links were emailed to the participant to complete in their own time in Multimedia Appendix 1. The survey was built using the Qualtrics survey tool [38]. Participants were asked to assess the software based on 10-point scales [36] and answer statements using a 5-point Likert [36,39-46]. The SUS was selected as it is suitable method when applied to a small sample size (N less than 14) [39]. Questions 1, 3, 5, 7, and 9 are positive, whereas questions 2, 4, 6, 8, and 10 are negative. The 10 connected questions provide a full review of a product. The SUS yields a score between 0 and 100 [47]. A higher SUS score is associated with greater product usability.

To evaluate the user interface, and identify problems with the software, heuristic evaluation was used [37]. There are several heuristic evaluation questionnaires used to assess human-computer interaction [37,48,49]. In this study, we used the heuristic evaluation questionnaire based on Sutcliffe and Gault's heuristic evaluation of VR apps [37]. It consists of 12 heuristic items, including natural engagement, compatibility with the user's tasks and domain, natural expression of action, close coordination of actions and representation, realistic feedback, faithful viewpoints, navigation and orientation support, clear entry and exit points, consistent departures, support for learning, clear turn-taking, and sense of presence.

Our survey was an open survey (no password required) based on several previous VR usability studies but modified slightly to align with our simulation [36,37]. The survey was checked by 10 individuals with a health care background to ensure it was easy to understand. In addition to the SUS questions and heuristic evaluations, we also collected information on demographics.

Procedure

At the beginning of the evaluation, participants were given a brief introduction to the project and shown how to use the VR headset and controllers. For those new to VR, an onboarding section was available. The overall evaluation ran for approximately 10 to 15 minutes. If there was more than 1 participant present at the same time, we allowed them to trail the software together so that they could see and interact with each other through the VRMDT. For those who evaluated solo, one of the development team would join them in the simulation so they could experience multiuser functionality. The participants were emailed the survey to complete within a 2 week time frame with a reminder sent after this period. Evaluations were conducted between February and March 2024). All sessions are located at The University of Manchester in a dedicated VR lab.

Data Interpretation

The results are interpreted as a grade for the SUS and a mean for the heuristic evaluation. To provide the grade of the SUS, there are 4 ratings for SUS interoperation: excellent (score greater than 80), good (69 - 80.3), okay (score equal to 68), poor (51-68), and awful (less than 51) [36]. For the heuristic evaluation, each item was rated for severity using Nielsen scale (no problem=0, cosmetic problem=1, minor problem=2, major problem=3, and catastrophe=4), as shown in Table 1 [47]. Only completed questionnaires were included in the final results.

Table . Nielson severity rating [48].

Rating	Definition
Don't Agree	I do not agree that this is a usability problem at all (there are no problems with usability)
Cosmetic problem	Needs not to be fixed unless extra time is available on the project (if there is time, aesthetic issue that only has to be fixed).
Minor problem	Fixing this should be given low priority (a low priority for a minor usability problem).
Major problem	Important fix required that should be given high priority (major usability problems, must be fixed right away)
Catastrophic	Imperative to fix this before product can be released.

Data Analysis

The final data were analyzed by entraining it into an Excel spreadsheet where the SUS score was calculated and the rate of the severity of each heuristic item based on the Nielsen severity scale for each item. The SUS questionnaire consisted of 10 questions. The score of SUS was calculated by adding the odd questions minus 5 and 25 minus the even number then multiplied by 2.5 [36]. On the other hand, the rate of heuristic severity was calculated by adding the number of statements and accepting the first statement which is no problem because it has zero value [37].

Ethical Considerations

The main purpose of this study is an anonymized evaluation of the VRMDT software in terms of its usability and utility. Therefore, the University of Manchester web-based ethics tool and the School of Health Sciences ethics representative confirmed that ethical approval was not required for this study. Consent was obtained from all participants that required them to sign a consent form. Anonymized responses were securely saved using the Qualtrics database.

Results

Participants

A total of 12 participants from a variety of health care fields were recruited (8/12, 67% female; 4/12, 33% male) with half of the participants being between 20 and 30 years of age (6/12, 50%). Most of the volunteers had a doctorate degree (8/12, 67%), with 4 having experience in health informatics. Most of the participants had no experience using VR before the evaluation. The demographic characteristics of the respondents are shown in Table 2.

Table . Demographics characteristics of the respondents (N=12).

Characteristics	Values, n (%)
Sex	
Female	8 (67)
Male	4 (33)
Age group (years)	
20-30	6 (50)
30-40	5 (42)
50-60	1 (8)
Highest education level	
Master degree	8 (67)
Doctorate degree	4 (33)
Background	
Nursing	3 (25)
Radiologist	1 (8)
Health Informatics	4 (33)
Medicine	2 (17)
Biostatistics	2 (17)

Usability (SUS Questionnaires)

A total of 67% (n=8) of participants gave SUS scores greater than or equal to 68. Four (33%) of the participants scored "Poor" with the VRMDT, with the SUS score rate less than 62. The total mean score was 72.7, resulting in an overall "Good" rating. The SUS scores for the respondents are shown in Table 3.

Multimedia Appendix 2 presents the interpretation of the SUS. Based on the SUS items, the participants indicated that the software was easy to learn how to use, with a mean score of 4.1. The highest score was given to the item "I found the various functions in this software were well integrated (eg, whiteboard, and DICOM images)" with a mean score of 4.25. In contrast, the lowest score was given to the item "I thought there was too much inconsistency in this software" with a mean score of 1.5 where low scores are an indicator of better consistency.

Table . System Usability Scale (SUS) scores for respondents. Average=72.7 (Good).

Respondents	Results			
	X0[a]	Y0[b]	SUS	Grade[c]
1	19	15	85	A
2	17	17	85	A
3	17	12	72.5	B
4	12	19	77.5	B
5	16	9	62.5	D
6	7	14	52.5	D
7	16	19	87.5	A
8	14	14	70	B
9	14	9	57.5	D
10	11	10	52.5	D
11	18	16	85	A
12	18	16	85	A

[a]The total odd SUS questions–5.

[b]25–the total even SUS questions.

[c]Sum of X0 and Y0 × 2.5 (A=Excellent, B=Good, C=Okay, D=Poor, and F=Awful).

Heuristic Evaluation

The participants rated the severity of each heuristic item based on Nielsen severity scale. The results of these ratings are shown in Table 4. The value of the first severity scale "no problem" is zero, so it was not counted. We estimated the number and severity of reported problems for each item. For example, we received 3 statements that indicated the minor problems for the first item "natural engagement," 1 for the major problem, and 2 for the cosmetic problem. The total score was calculated by adding each heuristic item. All the items had a usability score of less than 12, with a mean score of less than 2. This indicated well-functioning software.

The summary rate is shown in Table 5. One of the respondents reported 32 problems and 3 indicated no problems at all based on 12 heuristic items.

Table . Heuristics evaluation for each item with Nielson severity rating.

Number of items of the heuristics	Nielsen severity rating						
	No problem (0)	Cosmetic problem (1)	Minor problem (2)	Major problem (3)	Catastrophe (4)	Total	Mean
1. Natural engagement	9	0	2	1	0	3	0.7
2. Compatibility with the user's task	6	2	2	2	0	6	1.5
3. Natural expression of action	6	2	3	0	1	6	1.5
4. Close coordination	8	2	0	1	1	4	1
5. Realistic feedback	87	2	1	2	0	5	1.2
6. Faithful viewpoint	10	0	2	0	0	2	0.5
7. Navigation and orientation support	10	1	0	1	0	2	0.5
8. Clear entry and exit point	9	1	1	1	0	3	0.7
9. Consistent departures	8	3	1	0	0	4	1
10. Support for learning	7	0	4	1	0	5	1.2
11. Clear turn	11	0	0	1	0	1	0.2
12. Sense of presence	8	1	2	1	0	4	1

Table . Heuristics evaluation with Nielson severity rating for each respondent (resp).

Number of items of the heuristics	Respondents scores												
	Resp.1	Resp.2	Resp.3	Resp.4	Resp.5	Resp.6	Resp.7	Resp.8	Resp.9	Resp.10	Resp.11	Resp.12	Total
1. Natural engagement	0	0	0	2	3	0	0	2	0	0	0	0	0.5
2. Compatibility with the user's task	0	2	0	0	2	3	0	3	0	1	1	0	1
3. Natural expression of action	0	1	0	0	4	2	0	2	0	2	1	0	0.7
4. Close coordination	1	0	0	0	4	0	1	3	0	0	0	0	0.4
5. Realistic feedback	0	0	0	2	3	0	1	3	0	0	1	0	0.8
6. Faithful viewpoint	2	0	0	0	0	0	0	2	0	0	0	0	0.3
7. Navigation and orientation support	0	0	0	1	0	0	0	3	0	0	0	0	0.3
8. Clear entry and exit point.	0	0	0	2	0	0	1	3	0	0	0	0	0.5
9. Consistent departures	0	1	0	1	0	0	0	2	0	1	0	0	0.4
10. Support for learning	2	0	0	2	0	0	0	3	0	2	2	0	1
11. Clear turn	0	0	0	0	0	0	0	3	0	0	0	0	0.2
12. Sense of presence	0	0	0	2	0	2	0	3	0	0	1	0	0.7
Total	5	4	0	12	8	7	3	32	0	6	6	0	0.6

Discussion

Principal Findings

The findings of this study provide valuable insight into the current usability and future improvements of VRMDT software. Previous research into VR meeting rooms indicates that they may be an efficient tool for improving communication during the planning of patient treatments [50]. Kirchgessner et al [51] illustrated that VR meeting rooms are more motivational than traditional technologies such as Zoom. Our work supported this

with participant comments mentioning that presenting DICOM images in both 2D and 3D formats made the VR meeting more efficient than standard videoconferences with, respondent (D) mentioning “Being able to view images in 3D is the best thing about the VR software.”

Our results found that the VRMDT software had adequate usability, with a mean SUS of 72.7, which is classed as “Good” as an overall interoperation. Most of the participants indicated that the simulation does not require intensive training to use it, suggesting that the inbuilt onboarding software is sufficient for training purposes, the respondent (C) said that “Browsing menus was simple and they were easy to use. Viewing DICOM images was intuitive.” This is important for any health care institution as it will reduce the impact on existing training budgets and trainer time. Additionally, most of the respondents indicated that the software contains several useful functions, such as 2D and 3D DICOM views, a whiteboard, and an avatar that responds naturally to questions. These results suggest our software has clear advantages compared with conventional teleconferences. Another positive feedback was that the immersive 3D meeting room environment helped users feel as though they were in a real-world meeting. It is worth mentioning that a low score (mean=1.5) was given to the item “I thought there was too much inconsistency in this software,” which indicated that the software was more relevant to its aim and objectives, and it performed well. The heuristic evaluation method indicated that the VRMDT has a good user interface with a low number of reported issues.

User Experience

Participant feedback highlighted a few areas for improvement. Respondent (A) illustrated that “The reason why I indicated there were some problems was due to the internet connection not being stable, which sometimes led to lagging and the AI avatar being slow to respond,” and another respondent (B) said that “Software has potential but requires good Wi-Fi connection.” Therefore, one of the major issues indicated by most of the users was the poor internet connection, which effected the sense of presence and interaction with some functionality. Additionally, the internet connection effected the interaction with the avatar which resulted in delayed responses to questions. This was an issue with the evaluation room which received a poor internet signal and was not an issue with the software. The other issue was related to the avatar. The respondents mentioned that the AI needed to be further developed to respond to more specific clinical questions other than age, general treatment, and health conditions. Additionally, it should be designed to respond to any questions with different accent words, the respondent (C) said that “It also struggled with my accent for certain words.”

On the other hand, most of the respondents indicated that VRMDT was a powerful tool for sharing knowledge digitally compared with the other mediums because it contains several functions that make the environment immersive and very close to reality. Respondent (A) said that “it felt very futuristic, and I feel it will play an important role in future trans-geographical meetings.” Therefore, this software would be a good alternative tool in the future when face-to-face communication is not possible. Additionally, it was suggested that VRMDT may be an alternative tool for training and assessing the knowledge of junior professionals instead of in-person training. In the future, I would like to update the software by adding several functions that help in upgrading the current software. For instance, the meeting room will be secure under each hospital’s policies. In addition, those who have permission to enter this room can join this meeting after the invitation occurs. Moreover, It will contain the digital library, which contains the files and information about the cases that you want to make decisions regarding those cases.

Overall, the simulation was identified as a powerful tool for VR clinical meetings. In particular, it contained a functionality that allowed users to view both 3D and 2D DICOM images. While this has also been developed for off-the-shelf software (eg, [52]), the other software does not cater to a larger number of users generally seen at clinical meetings and lacks additional functionality such as a whiteboard, laser pointer, and AI-assisted avatar. Indeed, the avatar as an AI assistant was generally found to be very helpful in answering questions regarding the patients’ condition and was found to elevate the usability of the VR meeting. Previous independent work has suggested that cybersickness is an issue for some users [29,30,53]. That issue was not indicated in the user’s feedback from our study. The reasons for cybersickness not being an issue may include that the simulation was developed so the user can remain seated, which reduces excessive body movement both in real life and the simulation and provides a comfortable body position. Second, the headsets were modern (Meta Quest 3’s with battery strap) and had a high frame rate (90 Hz), with a wide field of view (110°H × 96°V), which also helps reduce the risk of sickness. The Quest 3 headset is also reasonably priced (£480; US $596) and easy to set up and use, making it a cost-scalable solution. We also found that the software was usable in the Meta Quest 2 without significant loss in performance, with this headset being a much cheaper option (£200; US $249.45). Overall, the hardware experience was good, with users finding the headset very light on their head, and the controllers easy to use. As a first-time exposure to VR, the majority found the experience “amazing” enough that they recommended its implementation for future VR meetings.

Limitation and Future Studies

This study has several limitations that are worth documenting, and which we will consider for future developments. First, the VRMDT software was evaluated by a small number of health care professionals. Second, most of the volunteers were researchers, and many were from the health informatics field. Third, we encountered another issue that the evaluation took place in a room that had a poor internet connection. That limited the testing of the software efficiently, particularly the avatar generative AI NLP which had lag, and multiuser functionality where verbal communication between users was slightly delayed. Finally, the generative AI seemed limited in answering questions related to the patient’s condition due to the lack of information available on the archival system.

Future research will need to consider testing using a more statistically powerful number of health care professionals

involved in MDT meetings to determine how powerful the 2D or 3D DICOM images are at identifying cancer lesions. Second, to overcome the internet issue, we need to test the network stability before performing the usability study. Thirdly, the AI generative avatar needs to be supplied with more detailed knowledge about the patients so it can more accurately answer. Additionally, a longitudinal analysis after implementation would allow researchers to assess the impact of the software on productivity. Finally, a direct comparison of our software with current digital tools such as Zoom and Microsoft Teams will help to assess its usefulness in terms of features, and productivity.

Conclusions

In health care institutions, applying knowledge management is crucial to using resources in a good way to increase patients' outcomes, and reduce medical errors. Knowledge sharing is considered an important step for the successful implementation of knowledge management. There are several factors that affect knowledge sharing in medical imaging. These factors can be divided into 3 categories: individual, departmental, and technological factors. MDT meetings are considered a crucial departmental factor in enhancing knowledge sharing. However, time constraints and geographical barriers can impact knowledge exchange efficiency. We have shown that creating a VRMDT meeting room may be a powerful tool to reduce those barriers.

Our VRMDT allowed the volunteers to interact with other users, and use the specialized features that allowed them to understand the patient's condition and scans in a correct and efficient way with the volunteers rating the simulation as good. Our results suggest that multiuser VR meeting rooms that use generative AI, and the ability to visualize DICOM files in both 2D and 3D have advantages over currently used meeting methods and would benefit from further development and research.

Future development and research by our group would evaluate the usability with a wider range of health care staff and an increased number of volunteers, and overcome the limitations that were outlined in this study. We also intend to explore software security for connecting to health care systems in order to access patient scans and data and develop the software across platforms to include a wider range of VR headsets as well as PCs and tablets.

Acknowledgments

JA is funded by the National Institute for Health and Care Research (NIHR) Manchester Biomedical Research Center. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health. This study was completed as part of doctoral studies funded by the Ministry of Health, Kuwait, and the University of Manchester, United Kingdom.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The consent form and the questionnaire of the survey.
[DOCX File, 28 KB - xr_v2i1e60651_app1.docx]

Multimedia Appendix 2
The interpretation of the System Usability Scale and heuristic evaluation.
[XLSX File, 14 KB - xr_v2i1e60651_app2.xlsx]

Checklist 1
STROBE Checklist.
[DOCX File, 35 KB - xr_v2i1e60651_app3.docx]

References

1. Sidpra J, Chhabda S, Gaier C, Alwis A, Kumar N, Mankad K. Virtual multidisciplinary team meetings in the age of COVID-19: an effective and pragmatic alternative. Quant Imaging Med Surg 2020 Jun;10(6):1204-1207. [doi: 10.21037/qims-20-638] [Medline: 32550130]
2. Singh RP, Javaid M, Kataria R, Tyagi M, Haleem A, Suman R. Significant applications of virtual reality for COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):661-664. [doi: 10.1016/j.dsx.2020.05.011] [Medline: 32438329]
3. Sadeghi AH, Wahadat AR, Dereci A, et al. Remote multidisciplinary heart team meetings in immersive virtual reality: a first experience during the COVID-19 pandemic. BMJ Innov 2021 Apr;7(2):311-315. [doi: 10.1136/bmjinnov-2021-000662] [Medline: 34192017]
4. Guo Z, D'Ambra J, Turner T, Zhang H. Improving the effectiveness of virtual teams: a comparison of video-conferencing and face-to-face communication in China. IEEE Trans Profess Commun 2009;52(1):1-16. [doi: 10.1109/TPC.2008.2012284]

5. Oeppen RS, Shaw G, Brennan PA. Human factors recognition at virtual meetings and video conferencing: how to get the best performance from yourself and others. Br J Oral Maxillofac Surg 2020 Jul;58(6):643-646. [doi: 10.1016/j.bjoms.2020.04.046] [Medline: 32417017]
6. Connelly M, Beals A, VanDerlip A, Singaraju K, Weintraub RL. Professional virtual communities for health care implementers: impact of participation on practice. iProc 2015;1(1):e11. [doi: 10.2196/iproc.4695]
7. Zhi H, Zhang D. Effects of professional virtual community attributes and flow experience on information sharing among Chinese young netizens: a fuzzy-set QCA. Psychol Res Behav Manag 2022;15:3497-3511. [doi: 10.2147/PRBM.S387166] [Medline: 36505670]
8. Munro AJ, Swartzman S. What is a virtual multidisciplinary team (vMDT)? Br J Cancer 2013 Jun 25;108(12):2433-2441. [doi: 10.1038/bjc.2013.231] [Medline: 23756866]
9. Kesson EM, Allardice GM, George WD, Burns HJG, Morrison DS. Effects of multidisciplinary team working on breast cancer survival: retrospective, comparative, interventional cohort study of 13 722 women. BMJ 2012 Apr 26;344:e2718. [doi: 10.1136/bmj.e2718] [Medline: 22539013]
10. Saini KS, Taylor C, Ramirez AJ, et al. Role of the multidisciplinary team in breast cancer management: results from a large international survey involving 39 countries. Ann Oncol 2012 Apr;23(4):853-859. [doi: 10.1093/annonc/mdr352] [Medline: 21821551]
11. Bydder S, Nowak A, Marion K, Phillips M, Atun R. The impact of case discussion at a multidisciplinary team meeting on the treatment and survival of patients with inoperable non-small cell lung cancer. Intern Med J 2009 Dec;39(12):838-841. [doi: 10.1111/j.1445-5994.2009.02019.x] [Medline: 20233244]
12. Friedland PL, Bozic B, Dewar J, Kuan R, Meyer C, Phillips M. Impact of multidisciplinary team management in head and neck cancer patients. Br J Cancer 2011 Apr 12;104(8):1246-1248. [doi: 10.1038/bjc.2011.92] [Medline: 21448166]
13. Taberna M, Gil Moncayo F, Jané-Salas E, et al. The multidisciplinary team (MDT) approach and quality of care. Front Oncol 2020;10:85. [doi: 10.3389/fonc.2020.00085]
14. Acai A, Sonnadara RR, O'Neill TA. Getting with the times: a narrative review of the literature on group decision making in virtual environments and implications for promotions committees. Perspect Med Educ 2018 Jun;7(3):147-155. [doi: 10.1007/s40037-018-0434-9] [Medline: 29797289]
15. Alexander T, Pfendler C, Thun J, Kleiber M. The influence of the modality of telecooperation on performance and workload. Work 2012;41 Suppl 1:3476-3483. [doi: 10.3233/WOR-2012-0627-3476] [Medline: 22317250]
16. Yunanda R, Cahyadi IJ, Theophilllus B, Oei J, Edbert IS, Aulia A. A systematic literature review: limitation of video conference. EMACS Journal 2022;4(3):119-122. [doi: 10.21512/emacsjournal.v4i3.8635]
17. González MA, Santos BSN, Vargas AR, Martín-Gutiérrez J, Orihuela AR. Virtual worlds. opportunities and challenges in the 21st century. Procedia Comput Sci 2013;25:330-337. [doi: 10.1016/j.procs.2013.11.039]
18. Abbas JR, O'Connor A, Ganapathy E, et al. What is virtual reality? A healthcare-focused systematic review of definitions. Health Policy Technol 2023. [doi: 10.1016/j.hlpt.2023.100741]
19. Baus O, Bouchard S. Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review. Front Hum Neurosci 2014;8:112. [doi: 10.3389/fnhum.2014.00112] [Medline: 24624073]
20. Vailshery L. Forecast for the number of active virtual reality users worldwide from 2014 to 2018 (in millions). : Statista; 2021.
21. Yamamoto GT, Altun D. Virtual reality (VR) technology in the future of military training. Presented at: 6th Istanbul Security Conference (2020); Nov 5-6, 2020; Istanbul.
22. Kyaw BM, Saxena N, Posadzki P, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959. [doi: 10.2196/12959] [Medline: 30668519]
23. Mazurek J, Kiper P, Cieślik B, et al. Virtual reality in medicine: a brief overview and future research directions. Hum Mov 2019;20(3):16-22. [doi: 10.5114/hm.2019.83529]
24. Almashmoum M, Cunningham J, Alkhaldi O, Anisworth J. Factors that affect knowledge-sharing behaviors in medical imaging departments in cancer centers: systematic review. JMIR Hum Factors 2023;10:e44327. [doi: 10.2196/44327]
25. Rankin NM, Lai M, Miller D, et al. Cancer multidisciplinary team meetings in practice: results from a multi‐institutional quantitative survey and implications for policy change. Asia-Pac J Clncl Oncol 2018;14(1):74-83. [doi: 10.1111/ajco.12765]
26. Streamlining Multi-Disciplinary Team Meetings: Guidance for Cancer Alliances: NHS England and NHS Improvement; 2020.
27. Walsh J, Harrison JD, Young JM, Butow PN, Solomon MJ, Masya L. What are the current barriers to effective cancer care coordination? A qualitative study. BMC Health Serv Res 2010 May 20;10:1-9. [doi: 10.1186/1472-6963-10-132] [Medline: 20482884]
28. Baniasadi T, Ayyoubzadeh SM, Mohammadzadeh N. Challenges and practical considerations in applying virtual reality in medical education and treatment. Oman Med J 2020 May;35(3):e125. [doi: 10.5001/omj.2020.43] [Medline: 32489677]
29. Srivastava K, Das RC, Chaudhury S. Virtual reality applications in mental health: challenges and perspectives. Ind Psychiatry J 2014;23(2):83-85. [doi: 10.4103/0972-6748.151666] [Medline: 25788795]

30. Rizzo AA, Strickland D, Bouchard S. The challenge of using virtual reality in telerehabilitation. Telemed J E Health 2004;10(2):184-195. [doi: 10.1089/tmj.2004.10.184] [Medline: 15319048]
31. Sentira XR. URL: https://sentiraxr.com/ [accessed 2024-03-12]
32. Inworld. URL: https://inworld.ai/ [accessed 2024-12-01]
33. ElevenLabs. URL: https://elevenlabs.io [accessed 2024-12-01]
34. Photon. URL: https://www.photonengine.com/# [accessed 2024-12-01]
35. Bevan N, Carter J, Harker S. ISO 9241-11 revised: what have we learnt about usability since 1998? human-computer interaction: design and evaluation. Presented at: 17th International Conference, HCI International 2015; Aug 2-7, 2015; Los Angeles, CA, United States.
36. Brooke J. SUS-a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester B, McClelland I, editors. Usability Evaluation in Industry: CRC Press; 1996, Vol. 189:4-7.
37. Sutcliffe A, Gault B. Heuristic evaluation of virtual reality applications. Interact Comput 2004 Aug;16(4):831-849. [doi: 10.1016/j.intcom.2004.05.001]
38. Machado C. Qualtrics research suite in academic context. Revista Onis Ciencia 2012;1(2):32-53. [doi: 10.21555/top.v32i1.171]
39. Grier RA, Bangor A, Kortum P, Peres SC. The system usability scale: beyond standard usability testing. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2013;57(1). [doi: 10.1177/1541931213571042]
40. Wang W, Cheng J, Guo JLC. Usability of virtual reality application through the lens of the user community. Presented at: CHI '19; May 4-19, 2019; Glasgow, Scotland. [doi: 10.1145/3290607.3312816]
41. Zhang T, Booth R, Jean-Louis R, et al. A primer on usability assessment approaches for health-related applications of virtual reality. JMIR Serious Games 2020 Oct 28;8(4):e18153. [doi: 10.2196/18153] [Medline: 33112243]
42. Moore N, Yoo S, Poronnik P, Brown M, Ahmadpour N. Exploring user needs in the development of a virtual reality-based advanced life support training platform: exploratory usability study. JMIR Serious Games 2020 Aug 7;8(3):e20797. [doi: 10.2196/20797] [Medline: 32763877]
43. Langlet BS, Odegi D, Zandian M, Nolstam J, Södersten P, Bergh C. Virtual reality app for treating eating behavior in eating disorders: development and usability study. JMIR Serious Games 2021 Apr 13;9(2):e24998. [doi: 10.2196/24998] [Medline: 33847593]
44. Foronda C, Mehta E, Morales MJ, Cacchione M, Gattamorta KA, Cardenas M. Usability of a virtual reality application to educate family caregivers of children with asthma. Clin Simul Nurs 2023 Nov;84:101465. [doi: 10.1016/j.ecns.2023.101465]
45. Shen J, Xiang H, Luna J, et al. Virtual reality-based executive function rehabilitation system for children with traumatic brain injury: design and usability study. JMIR Serious Games 2020 Aug 25;8(3):e16947. [doi: 10.2196/16947] [Medline: 32447275]
46. Jonathan E, Primasari CH, Purnomo Sidhi TA, Wibisono YP, Setyohadi DB, Cininta M. User interface evaluation in Metvaerse Gamelan virtual reality using heuristics evaluation method. Jour Int Intl Sys Soft 2022;1(2):136. [doi: 10.26798/jiss.v1i2.756]
47. Jordan PW, Thomas B, McClelland IL, Weerdmeester B. Usability Evaluation in Industry: CRC press; 1996.
48. Nielsen J. Severity ratings for usability problems. Papers and Essays 1995;54:1-2.
49. Nielsen J. How to conduct a heuristic evaluation. Nielsen Norman Group 1995;1(1) [FREE Full text]
50. Klemm P, Kleyer A, Tascilar K, Schuster L, Meinderink T, Steiger F, et al. A virtual reality–based app to educate health care professionals and medical students about inflammatory arthritis: feasibility study. JMIR Serious Games 2021 Dec 13;9(2):2535455269. [doi: 10.2196/23835]
51. Kirchgessner E, Sothers M, Aravena V, Baloian N. The mind in virtual meetings: comparing VR and video conferencing environments through experiential impact assessment and EEG analysis. Presented at: 15th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2023); Nov 28-30, 2023; Riviera Maya, Mexico. [doi: 10.1007/978-3-031-48306-6_26]
52. Medical imaging XR. URL: https://www.medicalholodeck.com/en/ [accessed 2024-03-04]
53. Nichols S, Patel H. Health and safety implications of virtual reality: a review of empirical evidence. Appl Ergon 2002 May;33(3):251-271. [doi: 10.1016/s0003-6870(02)00020-0] [Medline: 12164509]

Abbreviations

AI: artificial intelligence
DICOM: Digital Imaging and Communications in Medicine
MDT: multidisciplinary team
NLP: natural language processing
SUS: System Usability Scale
VR: virtual reality
VRMDT: virtual reality multidisciplinary team

Edited by L Riedemann, T Leung; submitted 17.05.24; peer-reviewed by T Soukup; revised version received 02.12.24; accepted 06.12.24; published 14.02.25.

Please cite as:
Almashmoum M, Payton A, Johnstone E, Cunningham J, Ainsworth J
Understanding the Views of Health Care Professionals on the Usability and Utility of Virtual Reality Multidisciplinary Team Meetings: Usability and Utility Study
JMIR XR Spatial Comput 2025;2:e60651
URL: https://xr.jmir.org/2025/1/e60651
doi:10.2196/60651

Development and User Experiences of a Novel Virtual Reality Task for Poststroke Visuospatial Neglect: Exploratory Case Study

Andrew Danso[1], PhD; Patti Nijhuis[1], PhD; Alessandro Ansani[1], PhD; Martin Hartmann[1], PhD; Gulnara Minkkinen[2], MA; Geoff Luck[1], PhD; Joshua S Bamford[1], PhD; Sarah Faber[3,4], PhD; Kat R Agres[5], PhD; Solange Glasser[6], PhD; Teppo Särkämö[1,7], PhD; Rebekah Rousi[8], PhD; Marc R Thompson[1], PhD

[1]Centre of Excellence in Music, Mind, Body and Brain, University of Jyväskylä and University of Helsinki, Seminaarinkatu 15, Jyväskylän yliopisto, Jyväskylä and Helsinki, Finland

[2]Department of Music, University of Jyväskylä, Jyväskylä, Finland

[3]Simon Fraser University, Burnaby, Canada

[4]University of Toronto, Toronto, ON, Canada

[5]Centre for Music and Health, Yong Siew Toh Conservatory of Music, National University of Singapore, Singapore, Singapore

[6]Melbourne Conservatorium of Music, The University of Melbourne, Melbourne, Australia

[7]Cognitive Brain Research Unit, Department of Psychology, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[8]School of Marketing and Communication, Communication Studies, University of Vaasa, Vaasa, Finland

Corresponding Author:
Andrew Danso, PhD
Centre of Excellence in Music, Mind, Body and Brain, University of Jyväskylä and University of Helsinki, Seminaarinkatu 15, Jyväskylän yliopisto, Jyväskylä and Helsinki, Finland

Abstract

Background: Visuospatial neglect (VSN) affects spatial awareness, leading to functional and motor challenges. This case study explores virtual reality (VR) as a potential complementary tool for VSN rehabilitation.

Objective: Specifically, we aim to explore the initial experiences of patients and physiotherapists engaging with a novel protocol, using an audiovisual cue task to support VSN rehabilitation.

Methods: A preliminary VR task integrating audiovisual cues was co-designed with 2 physiotherapists. The task was then tested with 2 patients with VSN over 12 sessions. The intervention focused on engaging neglected spatial areas, with physiotherapists adapting the task to individual needs and monitoring responses.

Results: Initial testing with 2 trainee physiotherapists indicated high usability, engagement, and perceived safety. Two patients with VSN completed 12 VR sessions. For Patient A, completion times increased following the introduction of an audio cue, though modeling indicated a nonsignificant linear trend (β=0.08; *P*=.33) and a marginally significant downward curvature (β=−0.001; *P*=.08). In contrast, Patient B showed a significant linear decrease in completion times (β=−0.53; *P*=.009), with a quadratic trend indicating a performance minimum around session 10 (β=0.007; *P*=.04). Intraweek variability also decreased. Motor scores (Box and Block Test and 9-Hole Peg Test) remained stable, and subjective feedback indicated improved mobility confidence and positive task engagement.

Conclusions: Further research with larger cohorts is needed to confirm the VR task's utility and refine the intervention.

(JMIR XR Spatial Comput 2025;2:e72439) doi:10.2196/72439

KEYWORDS

virtual reality; visuospatial neglect; physiotherapy training; audiovisual cues; patient experience; case report

Introduction

Background

Following a stroke, approximately 30% of stroke survivors experience neglect [1]. Neglect is a neurological disorder that poses significant challenges for rehabilitating behavioral deficits, including motor functions and perceptual-cognitive impairments, such as spatial awareness. Among the various types of neglect, visuospatial neglect (VSN) stands out as a specific subset. This is characterized by a failure to attend to objects or events within a defined region of the visual field, commonly affecting the left side [2]. These deficits increase the risk of falling and contribute to caregiver burden [3]. Conventional rehabilitative interventions typically involve manual interactions between patients and therapists, which can be physically demanding for practitioners and patients, leading to disempowerment, boredom, and reduced motivation when activities lack autonomy or engagement [4].

The integration of technology-based modalities, such as serious games and virtual reality (VR), with conventional rehabilitative interventions has emerged as a promising approach to engage patients with poststroke neglect [5]. When used in conjunction with traditional rehabilitation, including physiotherapy interventions, these modalities offer the potential for a more motivating treatment experience. Despite the potential benefits of these approaches, the use of audiovisual cues within VR adjunctively with physiotherapy remains poorly understood in terms of the subjective experience of patients with VSN. Consequently, the objectives of this case study are (1) to develop a real-time VR-based physiotherapy training solution tailored for individuals with VSN and (2) explore how audiovisual cues may influence the performance and rehabilitation experience of physiotherapists and individuals living with VSN during interaction with the VR-based training solution across 12 sessions.

Visuospatial Neglect

VSN, a common cognitive deficit following a stroke, is characterized by persistent spatial inattention, often manifesting unilaterally [2,6]. Patients with VSN struggle to acknowledge or respond to visual stimuli presented on the side opposite to the damaged hemisphere, often behaving as if that side of their visual field does not exist [7,8].

VSN is typically associated with damage to the posterior-parietal cortex of the right hemisphere. However, recent lesion mapping studies suggest a high degree of variability regarding the anatomical basis for neglect [9] with the temporo-parietal cortex, frontal cortex [10], occipital cortex [11], cerebellum [12-14], and even subcortical regions [15] have been linked to neglect. Furthermore, it has been associated with disconnections in white matter tracts, such as the superior longitudinal, inferior longitudinal, and inferior fronto-occipital fasciculi [9]. These varied findings highlight the complexity of VSN's neuroanatomical correlates.

Neglect mainly affects higher-level spatial processing modalities, such as visual and auditory spatial processing [16]. However, the empirical relationship between visual and auditory tasks with neglect remains unclear [8]. As studies have simulated multisensory (typically audiovisual and tactile) training procedures, improvements have been observed after training that used temporally congruent audiovisual input [16,17]. Therefore, the exploration of multisensory and specifically audiovisual training procedures is warranted.

Physiotherapy and Patient-Centered Treatment

Poststroke rehabilitation programs commonly feature physiotherapy to address motor and sensory impairments [18]. Physiotherapy is a vital primary care service within formal health care systems, aiming to sustain optimal physical functioning through various nonpharmacological interventions, such as progressive exercises [19,20]. Previous research [21,22] indicates that task-specific repetitive practice is essential for attaining lasting improvements in motor learning and motor function.

Recent trends in physiotherapy further emphasize the importance of patient-centered treatment [23]. Patient-centric physiotherapy treatment involves physiotherapists providing support to empower patients by providing emotional and physical assistance, alleviating fears and anxiety, and involving family and friends in treatment and care plans when possible. Patient-centered support often takes various forms of communication, including both verbal and nonverbal methods, such as tactile interactions and patient education [24]. Given the multifaceted nature of conditions, such as VSN, where multiple aspects of perception and movement may be affected, patient-centered support is particularly pertinent, as it addresses a wide range of dimensions in the recovery process.

Since the reaching and grasping skills of these patients are often limited, physiotherapy programs targeted for poststroke neglect rehabilitation include grasping training. Grasping training for poststroke neglect aims to improve spatial representation ability, as well as an enhancement in reaching, interacting, and grasping skills toward the neglected area or environment [25]. This often takes the form of congruent visual scanning training and motor rehabilitation tasks [26,27]. Studies suggest grasping training using methods such as home-based programs and custom-developed VR simulations is beneficial for patients with poststroke neglect to develop reaching and grasping skills that can be tailored to individual needs [28], which improves their ability to grasp objects [29-31].

VR and Neglect Rehabilitation

VR has emerged as a promising technology to be used adjunctly with physiotherapy, aiming to influence physical behaviors and movements within immersive, computer-generated environments. Sensory-motor tasks in VR offer several distinct advantages for physiotherapists. VR provides a safe setting for patients to engage in realistic and repetitive movements, either as an adjunct to conventional physiotherapy or in tandem with it, under the real-time supervision of therapists [5]. Evidence from various studies suggests that VR can improve the frequency of motor tasks in poststroke rehabilitation by increasing practice intensity [32], improving hand function [33], and promoting neuroplastic changes [34].

Recent studies have demonstrated VR's effectiveness in various stages of VSN management, ranging from diagnosis [35] and assessment [36] to motivation and rehabilitation [37]. This is in part due to VR's capacity to create immersive and controllable training environments, enhancing patient engagement and motivation, potentially leading to better treatment adherence and outcomes [38]. For rehabilitation, several studies have used VR to simulate realistic grasping training through hand grasp motions, showing promising results [39,40]. The engaging, adaptable, and measurable aspects of VR thus prove it to be a promising tool for VSN rehabilitation.

VR and Audio-Tactile Cueing in Neglect Rehabilitation

In VSN rehabilitation, audio-tactile cues enhance the immersive effects of VR by directing attention toward the neglected space through multimodal sensory engagement, thereby promoting orientation and visual awareness on the affected side. Studies by Knobel et al [41] and Leitner and Hawelka [42] provide evidence that audio-tactile cueing in VR settings with patients with VSN can effectively improve patients' attentional

orientation and head movement toward stimuli, assisting them to overcome rightward orientation biases. VR interventions can provide a structured and repeatable therapeutic experience, aligning with neuropsychological approaches (eg, prism adaptation therapy, a rehabilitation technique involving the use of prism glasses to shift the visual field and correct for visual displacement). Prism adaptation therapy has been integrated into VR environments, leading to more effective rehabilitation outcomes [43] as well as visual scanning training [42,44]. Phasic alertness (the brief adaptive increase in arousal that occurs in anticipation of an upcoming warning stimulus, see eg given by Posner [45]) has also been shown to be positively influenced with audiovisual cueing, leading to improvements in the balance of visual attention in patients with neglect [46]. Auditory cues can trigger fast, automatic shifts in spatial attention, suggesting preservation of strong links between auditory and visual attention mechanisms in patients with neglect. Sustained long-term improvements have been found following intensive and prolonged multisensory audiovisual stimulation [47].

This case study addresses a critical gap in the co-design and iterative development of VR-based interventions tailored specifically for hand grasping training in patients with VSN. Unlike generalized VR applications in rehabilitation, this intervention was designed through interdisciplinary collaboration with physiotherapists to integrate audiovisual cueing within a hand-grasping task, offering a novel approach to VSN rehabilitation [48]. For instance, existing VR-based interventions [49] have primarily focused on perceptual training through visual scanning tasks or general attentional cueing, whereas the system developed here aims to incorporate elements of compensatory motor initiation, less commonly addressed in this context. Compensatory motor initiation refers to the use of alternative motor strategies, such as gaze shifts, to facilitate movement toward the neglected hemispace, particularly in patients with VSN who exhibit impaired initiation on the contralesional side [50]. Accordingly, this VR intervention distinctly explores the integration of physiotherapist-informed design components, such as adjustable audiovisual cueing and targeted hand-grasping tasks, to address compensatory motor initiation and spatial attention. The structured co-design process included iterative testing and refinement to align the intervention with patient-specific needs and therapeutic goals [51]. In addition, exploring how such tasks influence individual patient experiences over multiple sessions provides valuable insights for personalizing rehabilitation strategies, addressing a critical need for evidence in this domain [1,52]. Furthermore, understanding how such a task influences the experience of individual patients over a series of physiotherapy sessions is unknown. Therefore, the following research question directed the study:

RQ. What are the initial experiences of patients and physiotherapists using a custom-developed VR-based hand grasping training protocol?

Accordingly, the aims of this study were 2-fold: (1) to develop a solution using audiovisual cueing to be used during real-time physiotherapy training and (2) explore the initial experiences and perceptions of patients and physiotherapists regarding the use of audiovisual cueing in the VR task.

Methods

Intervention Development

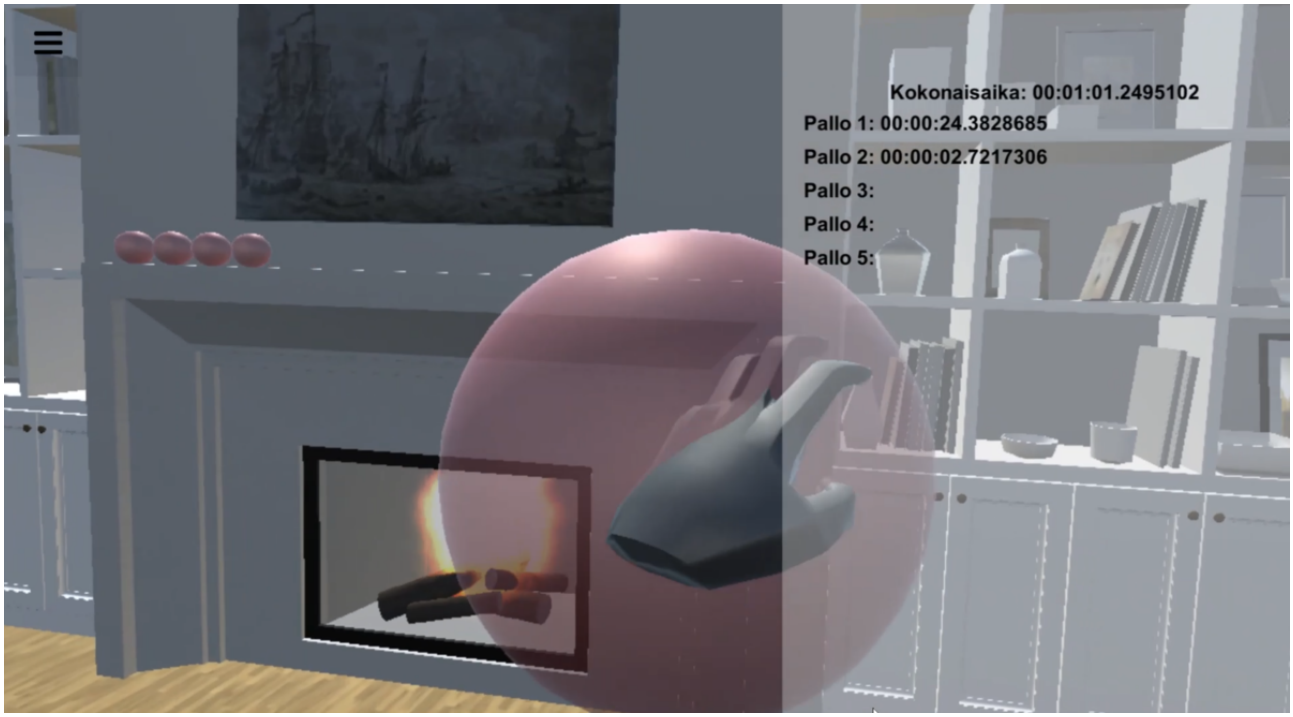
The case study presents a VR-based physiotherapy intervention designed for hand grasping training in the rehabilitation of VSN. This intervention uses VR to produce customizable visual and audio cues in its environment, aiming to address the requirements of individual patients with VSN. For instance, elements such as the timing, location, and dynamics of these cues can be adjusted to optimize the patient's training experience. In this intervention, participants engage in a VR task. The VR task is a single trial where a ball, serving as the visual cue, appears to the left within the VR environment, preceded by an audio cue to signal its location. The ball bounces in a fixed vertical up-down trajectory. This allowed users to plan their motor responses, such as grasping. Participants were tasked with grasping the ball as quickly as possible and were limited to grasping 1 ball per trial. In this study, we first focused on usability, followed by user testing. Audio cues were introduced in the seventh week to streamline the complexity of the VR task. This procedure-based approach to VR intervention design involves a structured process of usability testing and the phased introduction of multisensory cues to enhance task performance, aligning with principles of effective VR rehabilitation training [48].

Aligned with the UK Medical Research Council's (MRC) guidelines for developing complex interventions [51,53], this case study emphasizes a structured and evidence-based approach to intervention development. The UK MRC's emphasis on exploring feasibility and acceptability in the early stages of intervention development was addressed through detailed user testing (by both physiotherapists and patients with VSN) and design, to refine the VR task and ensure it met necessary user requirements.

Design and Implementation

For the purposes of this case study, an experimental VR environment was developed in Unity3D (Unity Software Inc). As shown in Figure 1, the VR environment contained multiple objects. Using Unity3D, the software developer crafted an experimental VR environment that featured a prominent visual cue in the form of a red ball. In addition, an auditory directional cue was used. The audio cue directed the user to the ball's location, and the task was considered accomplished upon successfully grasping the ball.

Figure 1. First-person perspective of the virtual reality task environment.



Directional Cue Design

Visual Cue

Figure 1 illustrates a red ball designated as the focal point for participant interaction, serving as a visual cue to direct their attention toward the task of grasping during trial runs. The figure depicts a VR task environment as viewed from a first-person perspective. In the center of the room, a red ball serves as the primary visual cue. The user's left hand is shown reaching for the red ball, indicating the action-based element of the task. The ball appears in the center, as this was captured during early development and captured only for example purposes. The interface also includes a timer display with labels in the Finnish language, such as "kokonaisaika," meaning "total time," and "Pallo 1" through "Pallo 5," meaning "ball 1" through "ball 5," respectively. These labels are followed by time stamps, revealing the duration taken to interact with each ball; for instance, "Pallo 1" took 24.38 seconds, and "Pallo 2" took 2.72 seconds. The sequence suggests that the user will engage with a series of 5 such balls throughout the exercise. The surrounding environment is minimally designed with a neutral color palette, emphasizing focus on the task elements. The Unity 3D rendered scene of the VR environment is presented in Figure S1 of the Multimedia Appendix 1

Developed using Unity 3D, the designated grasping zone is active from the ball's periphery to its center and is detected by Unity's collision system. The ball is designed to move vertically within the space (ie, vertical cues have been found to influence spatial orientation and potentially aid in rehabilitation of patients with VSN, see Lafitte et al [54]). The dynamics of this movement, such as the bounce speed, can be adjusted from 1 millisecond to 60 seconds through the application settings.

For this cue, a range of available adjustments was deemed important, as Golay [55] suggests that the effectiveness of cues in neglect rehabilitation can vary depending on the interval between the cue and the target. The starting position of the red ball correlates with the user's spatial location in the VR environment, which is determined by the positioning of their head-mounted display (HMD). The ball ceases movement when the participant's hand is near, simulating interaction. Successful grasping is indicated by the ball's disappearance. The appearance of the ball is designed to occur within the participant's left visual field (based on the [56] reach task consistently using targets appearing outside a central fixation point), determined by the spatial audio cues' effect on the participant's orientation in the HMD. The red ball appears approximately 61 centimeters (2 feet) from the participant, facilitating reach and interaction (eg, [57]).

Audio Cue

The audio cue was made by using the spatial sound capabilities of Unity3D 5.3, with the spatial blend parameter set to full 3D, allowing for precise auditory localization in conjunction with visual elements. Unity's spatial audio geometrically simulates sound sources within the environment, with the auditory cues emanating from the expected ball appearance location relative to the user's HMD position, facilitated by a head-related transfer function (HRTF) system. HRTF technology mimics how sound is affected by the listener's head and ears, providing a naturalistic sound perception based on directionality. The audio cue lasted 2.61 seconds and served to alert users to the specific location where visual stimuli would appear. This was based on prior research by Yoshizawa et al [58] demonstrating that a cue lasting 2 - 3 seconds effectively directed attention toward the neglected side during VR rehabilitation tasks for patients with hemispatial neglect. Studies by Dozio et al [59] and Knobel et al [41] suggest that short-duration audio cues are both beneficial and suitable in VR interventions for VSN rehabilitation. An interstimulus interval of 105 milliseconds between the auditory

and subsequent visual cue was optimized to prepare the patients for grasping the red ball (visual cue). The frequency spectrum of the auditory cue showed a prominent peak at approximately 500 Hz (≈ −30 dB) within the 50–20,000 Hz range, with additional smaller peaks across the low-frequency range. Figure 2 illustrates the frequency spectrum of the auditory cue. Figure 3 provides a 3D visualization of the auditory cue.

Figure 2. Frequency spectrum plot of auditory cue.

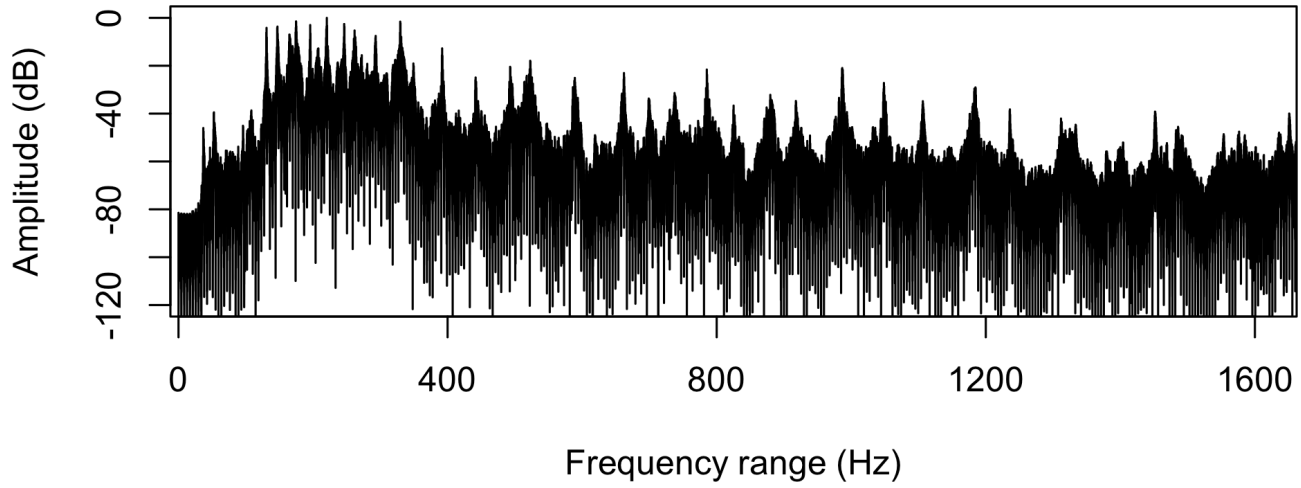
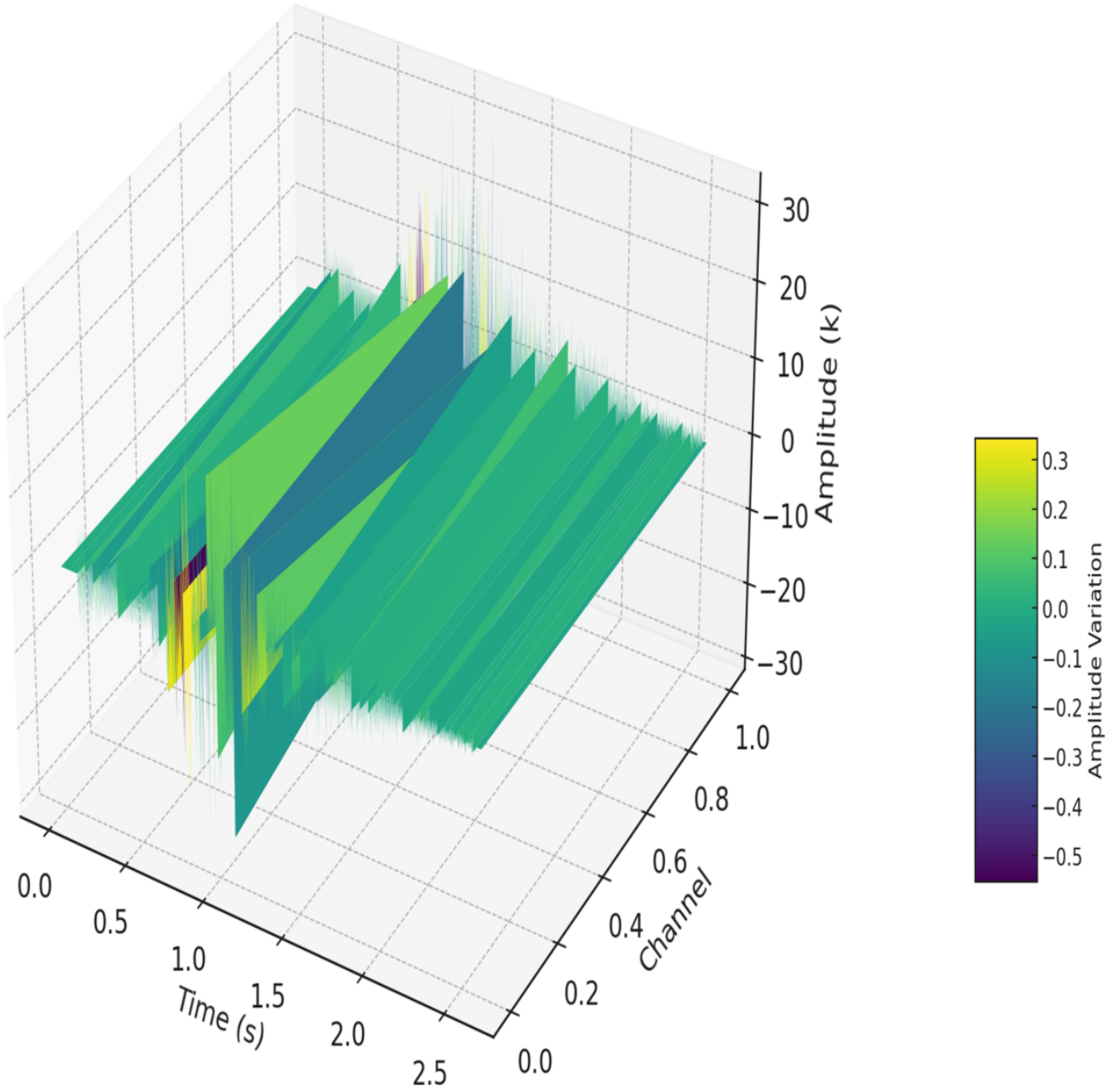


Figure 3. 3D visualization of auditory cue. The x-axis—Time (s)—denotes time in seconds, the y-axis—Channel—displays the left and right audio channels, and the z-axis—Amplitude (k)—illustrates the amplitude scaled in kilounits. The color gradient, as indicated by the color bar labeled "Amplitude Variation," visually depicts the amplitude fluctuations within the auditory cue.



Task Parameters and Environment Design

VR Task Description

The VR task includes 1 trial with a ball appearing to the left (15° to the left), within a 30° horizontal plane and a 50° vertical plane within the limits of the VR room. To complete each trial, participants are required to successfully grasp the floating ball as fast as possible (includes a 5-minute time-out period for managing patient fatigue, [22,23,60]). Upon appearing, the ball bounces with a vertical up-down trajectory in the room and stays bouncing within a fixed vertical trajectory until the participant grasps the ball. Participants were limited to grasping 1 ball once per trial. To initiate the appearance of a new ball, participants were required to rotate their trunk and direct their gaze toward the center point of the field of view (FOV). This central gaze point was represented in the VR software as a painting positioned above a fireplace object. The process was marked by a countdown timer, starting from 3 and concluding at 0, at which point a new ball was generated. The duration of each trial was measured (in ms) until a successful grasp occurred. The ball's visual stimuli were depicted through its appearance in the VR environment. An audio cue of where the ball will appear across the 30° horizontal plane was activated prior to the ball appearing to alert the patient to the appearance location (interstimulus interval between the audio cue and visual cue=105 ms, ie, [41]). There were 15 trials in each location across the 30° x 50° degree plane. The timer's data aimed to provide insights into how target grasping efficiency is affected by the size and distance of objects. In addition, multiple trials were incorporated and the elimination of manual restarts to improve the interaction process and reduce the cognitive load associated with initiating new trials. Data output was presented

in the form of a text file, which recorded participant response times for each trial, audio cue sounds, and trial dates.

Task Development Aims

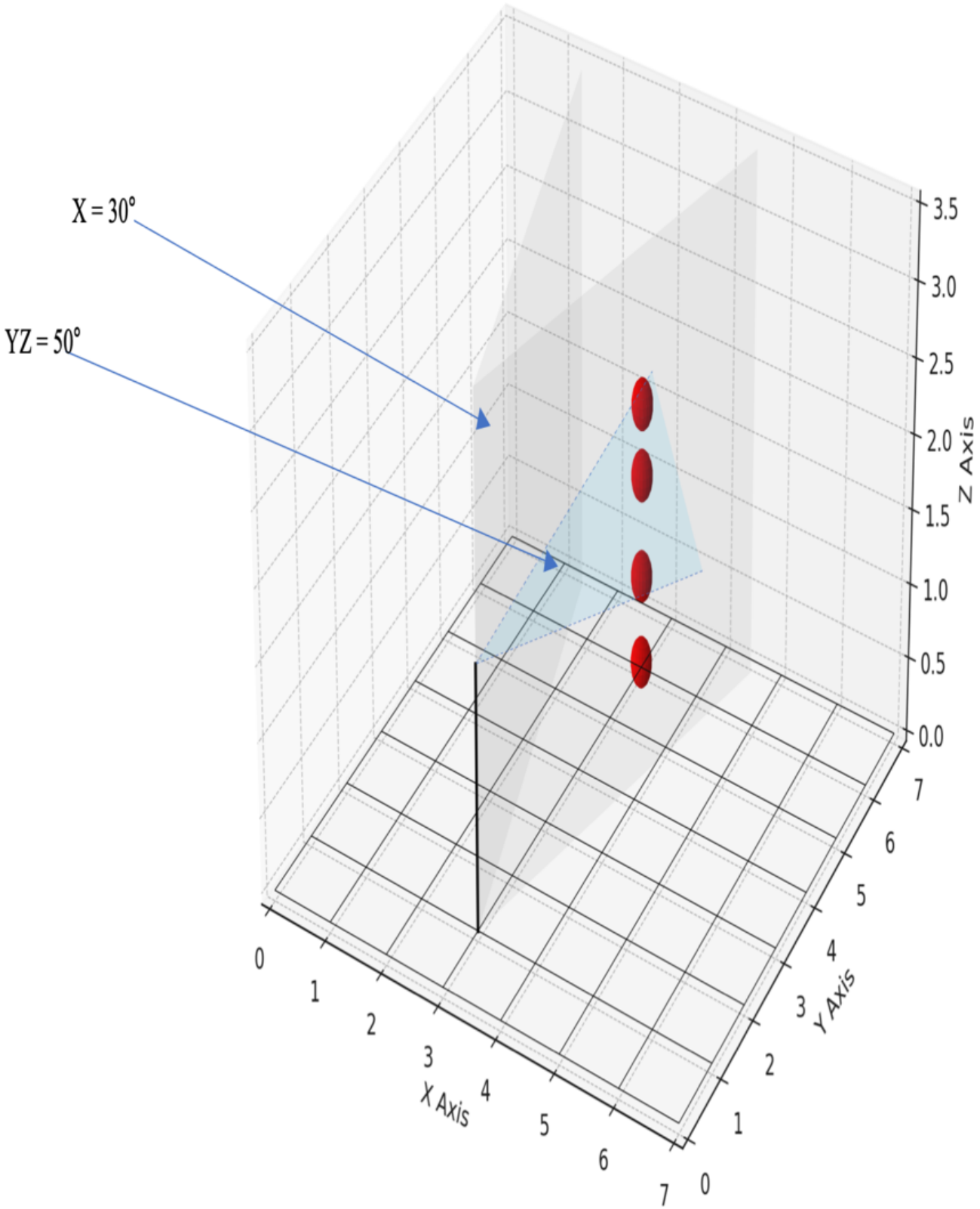
The development of this VR task is specifically tailored for inclusion in poststroke physiotherapy rehabilitation, building upon the Reach Task conceptualized by Mattingley et al [56]. The original task involved participants reaching toward a stimulus at the edge of their visual field. Our adaptation for VR purposes follows this principle, focusing on the neglected visual field to help distinguish perceptual deficits from motor control difficulties.

To support physiotherapeutic goals, the task encourages the use of compensatory strategies—alternative motor patterns developed to adjust for lost function, as described by Levin et al [61]. Patients with VSN engaging in these motor patterns is critical for fostering attentional shifts [62]. Integrating audio cues with visual targets is designed to enhance anticipatory behavior [55], supporting patients with neglect to proactively direct their attention and gaze toward the task at hand.

Applying Fitts Law as a Guiding Principle for Task Difficulty

Fitts law is a psychological principle stating that the difficulty of a perceptual-motor task, such as pointing or selecting targets, is a function of target size and distance [63]. The smaller the target size, the slower and more difficult it is for individuals to accurately reach or activate the target. This provided a framework for reducing the difficulty by enlarging the target and thereby increasing the accessibility of user interactions within the VR environment. As part of subsequent development tasks, the software developer adjusted the FOV to 30° on the horizontal plane and 50° on the vertical plane using Unity's built-in parameters. This adjustment was chosen to reduce the distance to targets, thereby making the task more accessible for patients with VSN to successfully grasp objects (this also reduced the potential for cybersickness effects, eg, [60]). The adjustment of the FOV to specific angles on the horizontal and vertical planes is also made in relation to the effective size and position of the targets (ie, the balls) within the VR environment (see Figure 4 depicting a conceptual diagram representing the appearance location for the primary visual cue within the updated FOV parameters). The ball was set to appear at ground level, ascend vertically to the ceiling, and descend vertically back to its initial point of origin on the ground (up-down trajectory).

Figure 4. Illustrative figure depicting the field of view in the virtual reality task. This figure depicts the field of view in the virtual reality task, with 4 red spheres indicating potential appearance locations of the primary visual cue within a 15° range to the left of the central gaze. The black line represents the participant's position at coordinates x=3, y=1. “X=30°” and “YZ=50°” denote the maximum horizontal and vertical area visible to the participant. The 2 blue arrows indicate the breadth and height of the participant's potential visual engagement area during the activity. The diagram is conceptual and not drawn to scale; axis measurements of 7x7 meters (X and Y) and 3.5 meters (Z) are for reference only.



High Tech Computer Corporation Vive Head Mounted Display

The VR intervention was delivered using the High Tech Computer Corporation (HTC) Vive HMD, a tethered head-mounted display equipped with lighthouse tracking technology for accurate room-scale motion capture. Its compatibility with HRTF audio makes it suitable for delivering spatial audio cues, essential for the task design. A detailed specification of the VR hardware setup, including display resolution, connectivity, and audio components, is provided as supplementary information (Section A in Multimedia Appendix 1).

Leap Motion Controller

Hand tracking in the VR task was achieved using the Leap Motion Controller, a touchless optical tracking device that allows real-time monitoring of hand and finger movements [64]. This interface enabled intuitive grasping interactions without the need for handheld controllers. Haptic feedback was not included, in line with previous studies highlighting the complexity it introduces in poststroke rehabilitation tasks [64,65]. A specification of the Leap Motion Controller and how it pertains to this study is provided as supplementary information (Section B in Multimedia Appendix 1).

Usability and Preliminary Testing

Preliminary User Testing

During the initial development phase, collaboration took place between a physiotherapist from the anonymous physiotherapy clinic, who played the role of a user tester, and a software developer from (anonymous organization). For testing purposes, they used a PC-based system along with a tethered HTC Vive HMD and a Leap Motion Controller. Initially, challenges emerged related to latency, particularly concerning the responsiveness of the Leap Motion Controller device to the grasping gesture. To overcome this issue, they decided to externally mount the Leap Motion Controller on the HMD to enhance the tracking of hand movements and gestures, serving as a trigger point for task completion. Mounting Leap Motion Controllers on a VR HMD has been found to enlarge the tracking area for hand tracking in VR software programs, thereby improving user experience with enhanced hand and gesture tracking [66].

Usability Assessment

To assess usability as well as the physiotherapists' perception of the task and system, 4 questions from the Technology Acceptance Model questionnaire [67], 2 questions from the Engagement in In-Game Questionnaire [68] and 1 question from the Safety Attitudes Questionnaire [69] were adapted and used for a usability assessment. The questions were answered using a 5-point Likert scale, ranging from "fully disagree" to "fully agree" (see Table S1 in Multimedia Appendix 1).

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki [70] and a favorable ethical statement from the Pirkanmaa Ethics Committee (984/2021). Written informed consent was obtained from all participants. Personal data were processed under a Data Processing Agreement in compliance with the GDPR (EU, 2016/679). Data were pseudonymized, stored securely on password-protected servers, and accessible only to authorized researchers. No identifiable images or personal identifiers are included. Participants were reimbursed for travel expenses but received no other compensation.

Results

Feasibility and Usability Feedback (Physiotherapists)

Before conducting tests with participants with VSN, 2 trainee physiotherapists from an anonymous physiotherapy clinic participated in VR task test sessions to gain insights into their experiences with the technology. Subsequently, these test sessions were immediately followed by a survey where the 2 physiotherapists provided feedback regarding their subjective experiences with the VR task. The survey was administered using Webropol (Webropol Oy) software. Each trainee used the application for approximately 10 - 15 minutes before completing the survey. Table 1 aimed at assessing (1) the physiotherapists' perceived ease of use of the system, (2) the physiotherapists' engagement while using the system, and (3) their perception of the safety of the VR task. The questions for usability (Question 1–Question 4), task engagement (Question 5), and safety (Question 7) were rated highly by both trainee physiotherapists. The question associated with difficulties in learning the task (Question 6) was rated low.

Table . Questionnaire results from 2 trainee physiotherapists.

Survey question	Physiotherapist 1 rating	Physiotherapist 2 rating
I think the system is easy to use	4	4
Learning to use the system is not a problem	4	5
I enjoyed using the system	4	4
I would like to use the system in the future if I had the opportunity	4	4
Learning to use this VR[a] task was easy	4	4
Was learning the task difficult?	2	2
I would feel safe using this as a patient	4	4

[a]VR: virtual reality.

Following feedback from the physiotherapists during initial testing, session length was set to approximately 10 - 15 minutes. While fatigue was not directly measured, this duration was selected to support tolerability and aligns with findings recommending shorter VR sessions (10 - 30 min) to reduce fatigue in poststroke rehabilitation [71].

Patient Characteristics and Baseline Function

Patients were recruited between February and March 2022 from inpatients with poststroke neglect referred to care at the anonymous hospital/clinic, Finland. Inclusion criteria required participants to have a right hemisphere stroke with diagnosed neglect, be right-handed, be medically stable, be without hearing impairments, cognitive deficits (eg, learning difficulties), or hemianopia (loss of vision in one-half of the visual field), and be aged 18 years or older. In addition, patients were assessed for their physical and cognitive ability to perform the audiovisual VR task by a physiotherapist team. Patients were excluded if motor or communication impairments, as determined by the physiotherapists, were severe enough to prevent task participation or understanding of instructions. Eligible inpatients received detailed study information, and participation was discussed.

Based on the results of the previous phase of testing, and on the expertise of the physiotherapy clinic, we decided to use the VR task as part of physiotherapy sessions with 2 patients with poststroke neglect for the next phase of exploration. To preserve procedural integrity and ensure personalized care, sessions were conducted by 2 licensed physiotherapists from an anonymous hospital/clinic. Patient A was a 46-year-old male with left-sided hemiparesis and VSN, 1 year poststroke (Barthel Index: 70/100). Patient B, a 37-year-old female with hemiplegia and VSN, was 4 years poststroke (Barthel Index: 95/100). Both were right-handed and met inclusion criteria (full clinical profiles and ADL scores are presented in Table S1 and Section D in Multimedia Appendix 1).

The inclusion of 2 patients with differing clinical profiles, a mild case (Patient A's mild hemiparesis) and a more severe case (Patient B's severe hemiplegia), was a deliberate methodological choice consistent with early-phase intervention research (eg, UK MRC guidelines [51,53]). This heterogeneity enabled an initial assessment of the VR system's usability across a spectrum of functional symptoms and rehabilitation timelines. Such purposive sampling is supported in the development of complex interventions, where the goal is to evaluate feasibility, individual responsiveness, and context-specific implementation [51,53,72]. In neurorehabilitation, evidence shows that early inclusion of diverse patient profiles enhances understanding of task usability, supports iterative design, and informs future personalization strategies [1,41]. Furthermore, diverse case inclusion enables deeper insight into patient-centered customization [17,52] (see eg, [73] where VR task parameters were adapted to patients with varying upper-limb impairments, improving usability, and elevating future patient adherence to the intervention).

Description of Patient Test Sessions

Physiotherapy interactions were standardized across both patients to ensure procedural consistency while supporting individual needs. Both licensed physiotherapists underwent training in the VR task and applied identical task parameters (eg, 1.19 s ball bounce, consistent audio cue use), emphasizing procedural integrity and a patient-centered framework [74]. Patients received uniform instructions and completed preparatory sessions to familiarize themselves with the VR environment. During sessions, physiotherapists monitored performance in real time and provided feedback based on individual motor behavior, such as compensatory strategies (eg, trunk rotation or delayed reaching). Therapists manually initiated each trial using the in-task menu (see Figure S2 in Multimedia Appendix 1). A detailed description of session setup, training, and interaction procedures is included as supplementary information (Section C in Multimedia Appendix 1).

Assessment Measures

During this phase, the assessment measures encompassed the following: time to completion data (with successful grasps serving as indicators of task completion); initial rehabilitation goals set by the physiotherapists and patients prior to commencing the 12 sessions incorporating the VR task (see Textbox 1); an evaluation of goal attainment postcompletion of the 12 sessions involving the VR task; 2 standardized motor function assessments (ie, the Box and Block Test, BBT, which measures gross manual dexterity—number of blocks moved in 60 s—and 9-Hole Peg Test, 9HPT, which assesses fine motor coordination in seconds [75]; see Table 2); and each patient's subjective experience, documented through their comments following the completion of the 12 sessions.

Textbox 1. Textbox 1. Physiotherapy goals for patients A and B as reported by their physiotherapist.

Patient A goals

- Ability to move in an upright position
- Strengthening of leaning on the left side of the body

Patient B goals

- To gain confidence in walking
- Improve balance
- Muscle condition improvement

Table . Patient A and B's Box and Blocks Test and 9-Hole Peg Test scores.

Patient	Hand	BBT[a,b] Pre (blocks)	BBT Post (blocks)	9HPT[c,d] Pre (s)	9 HPT Post (s)
A	Left	27	27	21.30	22.40
A	Right	51	50	20.84	23.30
B	Left	23	22	20.77	21.50
B	Right	54	54	21.60	21.20

[a]BBT: Box and Blocks Test.

[b]Scores indicate the number of blocks a patient can move over a partition from one compartment to another within 60 seconds, using one hand.

[c]9HPT: Nine-Hole Peg Test.

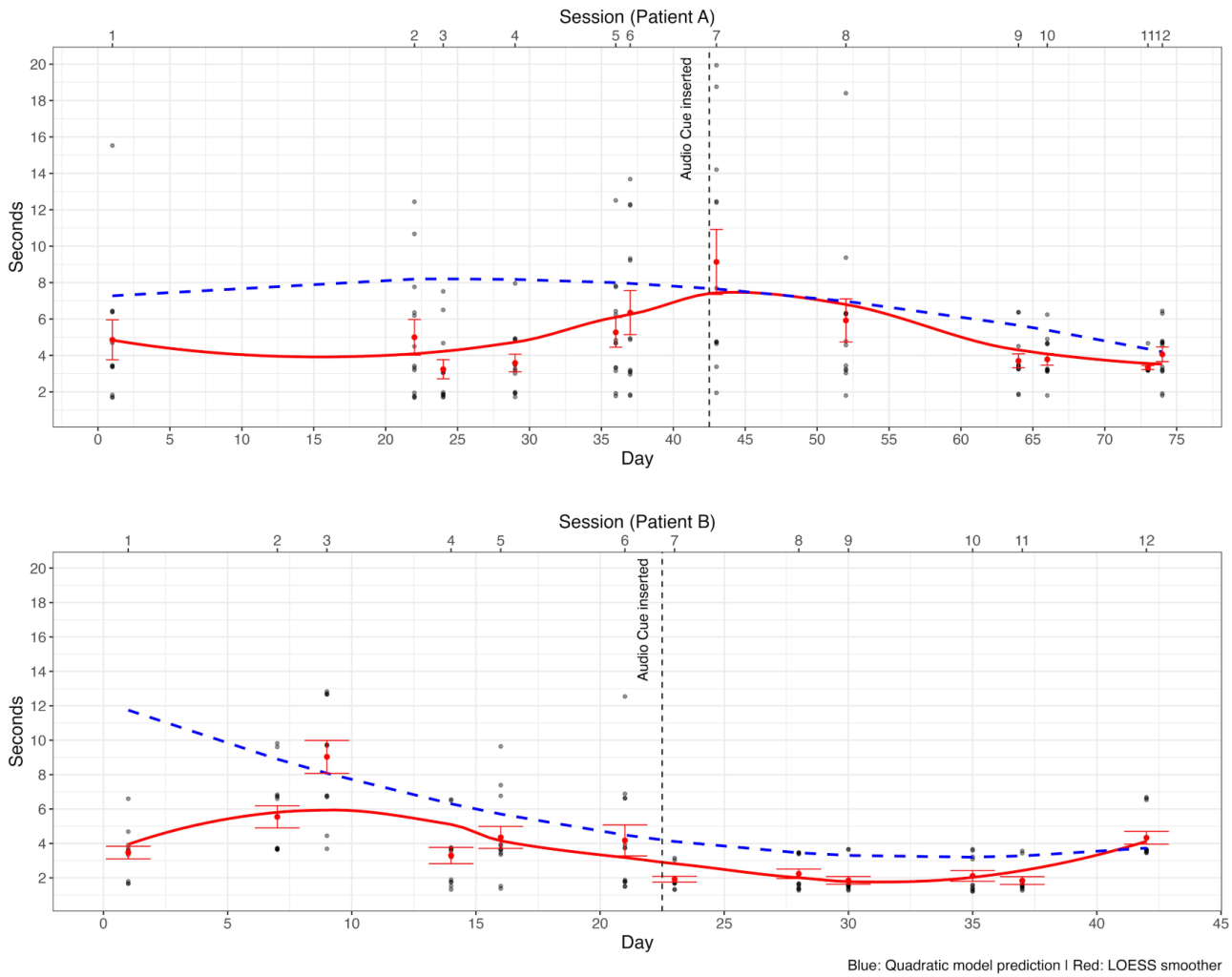
[d]Scores provide a standardized measure of fine motor dexterity, particularly assessing hand–eye coordination, finger function, and speed of movement during a precision-based task.

Throughout these initial sessions, both physiotherapists guided their patients to begin the task by focusing their visual attention above the fireplace in the VR environment, focusing on the painting object. Once each patient's gaze was visually focused, a countdown timer, counting down from 3 to 0, initiated the appearance of a ball within the FOV. The patients were then instructed to reach out and grasp the ball as it appeared. Upon the patients' successful grasping of the ball and subsequent completion of this first trial, both patients completed an additional 14 trials before the task ended. The physiotherapists would remain present with their patients throughout the entirety of this study to provide additional support or further instructions that might be needed by the patient. This was also due to safety reasons, as both patients interacted with the task while standing upright.

Patient Completion Time Description

We obtained 180 trials per participant. Before proceeding to data inspection, the data were trimmed to remove the worst (ie, slowest) 2 trials for each week. A visual inspection of the task completion times of both patients is presented in Figure 5. Conducting statistical analyses on 2 patients is supported within the framework of case study methodology, which allows for the test of intervention effects on an individual basis [76].

Figure 5. Patient task completion times. Dotted data points illustrate the completion times of each trial. Whiskers denote observed SE of the mean for each day, based on the trial-level data. The solid line signifies a loess interpolation, providing a continuous representation of the completion times over time. The dashed line indicates the trend predicted by the model. To ease visualization, dots above 20 seconds (NA=12; NB=7) are not displayed.



A curve estimation analysis was conducted to examine the trends of the completion times throughout the sessions (N=12 with 15 trials each) for both patients independently. Consistent with the positive skewness of the completion times (Skewness for Patient A=2.63; Skewness for Patient B=4.11), we resorted to a generalized linear model approach through R's *lme4* package [77]. The data were modeled using a gamma distribution, which is ideal for positively skewed strictly positive continuous data [78]. To facilitate interpretation, no link function was used, allowing coefficients to be interpreted directly on the original scale of the outcome. In greater detail, 6 curve models were fitted to the data (ie, linear, logarithmic, quadratic, power, inverse growth, and exponential decay), consistent with methodologies in rehabilitation research that use curve estimation to track patient progress over time [79]. Subsequently, the models were compared based on the Akaike Information Criterion, Bayesian Information Criterion, Nagelkerke R^2, and performance score (through the *performance* R package [80]) (see eg, [81,82] model comparison and information criteria). In all models, the day of the session was used as the predictor. This was preferred over the session number because our sessions were not equally distant in time. For both patients, the quadratic model showed the best fit (see Table 3).

Table . Model comparison.

Patient and model	AIC[a]	BIC[b]	R^2	Performance
Patient A				
linear	1022	1032	0.070	50.04
quadratic	1019	1032	0.104	100.00
inverse	1031	1041	0.013	0.39
log	1029	1039	0.027	7.63
exp	1031	1041	0.012	0.00
power	1031	1040	0.016	1.96
Patient B				
linear	941	950	0.146	47.02
quadratic	932[c]	945[c]	0.205[c]	85.40[c]
inverse	962	972	0.004	11.09
log	945	954	0.120	28.07
exp	962	972	0.002	14.87
power	957	966	0.042	3.50

[a]AIC: Akaike Information Criterion.

[b]BIC: Bayesian Information Criterion.

[c]Best models.

Regarding Patient A, when inspecting the raw data, a marked increase in completion time was found at session 7, namely, the session wherein the audio cue was added. However, the model suggested that, after a nonsignificant initial increasing trend (β_{linear}=0.08, SE=0.08; *P*=.33), a downward curvature began to emerge, approaching significance ($\beta_{quadratic}$=−0.001, SE=0.001; *P*=.08), around day 24 (ie, session number 3).

Model's results for Patient B indicate that the completion times exhibited characteristics of a significantly decreasing duration (β_{linear}=−0.53, SE=0.20; *P*=.009), where initial task completion times were followed by a gradual deceleration toward a minimum time for task completion, reached around day 35 (ie, session number 10; $\beta_{quadratic}$=0.007, SE=0.003; *P*=.04). Furthermore, Patient B also showed a strong reduction in average completion times and intraweek variability after session 6.

Assessment of Upper Limb Motor Function

The BBT and 9HPT were included as standardized measures of gross and fine motor function, respectively, to assess upper limb performance relevant to the grasping demands of the VR task.

Table 2 summarizes pre- and postintervention scores from the BBT and 9HPT, which were performed under therapist supervision, with patients instructed to direct their focus toward their neglected side. Results show consistent right-hand performance on the BBT while the left (affected) hand showed mild impairment.

Patients' Subjective Experience

At the conclusion of the 12-session period, Patient A and Patient B were individually asked to provide feedback regarding the use of the VR task in conjunction with their physiotherapy treatment (see Table 4 below for a summary of Patient A and B's qualitative feedback). These feedback rounds were conducted by their respective physiotherapists, each posing the same set of two questions: (1) To what extent were you able to achieve your rehabilitation goals? (2) How would you describe your experience with the VR task, using your own words?

Table . Qualitative feedback from Patients A and B in response to 2 questions posed by their physiotherapist.

Questions	Patient A feedback	Patient B feedback
Were your rehabilitation goals achieved?	"In my opinion, there is no difference in my condition."	"Shifting weight to the left at the end of the period was easier and was slightly more successful. During this period, I started walking without support, without a walking stick." "Walking has become more confident."
How did you experience the VR[a] task (in your own words)?	"The VR task was fun and interesting. The rehabilitation experience was different from my normal physiotherapy experience."	"It felt quite nice and relatively easy. When the balls were on the left *and* behind, it was more difficult. The audio did not seem to really help my performance in the game. But they calmed my thoughts down. The most challenging was when I couldn't move my legs along but, closer to the end it was easier."

[a]VR: virtual reality.

Both physiotherapists transcribed the responses provided by their patients, with the conversations taking place in the Finnish language.

Table 4 indicates that while Patient A did not consciously notice a difference in his condition, he experienced the VR task as fun and interesting. Patient B actively noticed a difference, potentially contributed to by engagement in the VR rehabilitation task. Patient B explained that the VR technique was partially more effective and that shifting weight was easier. In addition, the patient began to walk without support following the VR tasks. Patient B also explained the difficulty in grasping for the ball when it was in the area affected by neglect yet noted that in addition to being a relatively easy exercise, the audio calmed them.

Discussion

Principal Findings

The objectives of this case study were 2-fold: (1) to develop a VR-based solution incorporating audiovisual cueing, designed for real-time use during physiotherapy training sessions for poststroke VSN rehabilitation and (2) to explore the initial experiences and perceptions of both patients and physiotherapists regarding the use of audiovisual cueing within this VR task during rehabilitation. The development process resulted in the successful development of a perceptual motor task customized to meet the needs of real-time physiotherapy applied adjunctly to patients with VSN. Both patients reported a positive subjective experience with the VR training, citing enjoyment and interest, and 1 patient even experienced some improvement in motor function. The VR training was also positively received by physiotherapy trainees. Taken together, these results provide several promising case-specific elements and a potential roadmap for future task development, as well as a larger trial with explicit control and standardization.

Development Aims

The VR task incorporated a progressive approach to rehabilitation [19,20], as evident in the task design. For instance, several adjustments were made in response to user feedback (eg, accessibility considerations were integrated into the system, with a deliberate adjustment of the FOV to cater to patients with VSN). Notably, the usage of VR technology depended on the provision of training to the physiotherapists to use the technology prior to testing it with patients. This training was a vital aspect of the study, ensuring they possessed the necessary expertise to integrate technology into rehabilitation practices safely. Insights from the developer and shared experiences from physiotherapy trainees emphasize the collaborative nature of the approach, incorporating external perspectives and expertise.

Patient A and B Qualitative Feedback

As we used a patient-centered approach, the subjective experience of patients was a key outcome measure. Generally, the qualitative feedback from both patients highlights positive engagement with the task. Both patients reported positively about their experience, stating: "the VR task was fun and interesting" (Patient A) and the VR task was "quite nice" and "relatively easy" (Patient B). The VR task offers a different experience, which was positively received, as Patient A states: "the rehabilitation experience was different from my normal physiotherapy experience." This indicates benefits beyond strictly functional and medical outcomes, such as increased patient engagement during physiotherapy training, which may find alignment with [83].

Patient B's feedback also reflects progress toward achieving rehabilitation goals (eg, the objective of regaining confidence in walking). However, Patient A perceived no significant difference in their condition, which may be due to a multitude of factors. While Patient B also notes some challenges, suggesting potential difficulties related to spatial awareness or balance training, they state that with training, the task did become easier. This may indicate that the VR task was challenging the patient with positive training outcomes as a result. The patient's comment about the game's audio—"it provides a calming effect"—while not intended, aligns with literature on sensory stimulation having a calming effect on patients [84] Though not directly impacting performance, it implies the potential for sensory engagement as a therapeutic aid. However, the functional clarity and perceived usefulness of the audiovisual cues were not explicitly evaluated in this study, pointing to the need for future iterations to include cue-specific assessment. Indeed, as indicated by Danso et al [52], more research must be done to systematically study the impact of sound and music on therapeutic progress. Despite being at distinctly different stages of poststroke recovery and

presenting with varying symptomatic profiles, both patients reported a positive experience. This outcome highlights the value of customizing the VR task, including directional cues (audio and visual) and in-task settings (eg, background music volume, audio cue volume), to suit individual needs.

The continuous presence of the physiotherapists with their patient throughout the duration of the study, while using VR, suggests a patient-centric approach [24,85]. They actively assisted patients in fitting the HMD as well as manually initiating and monitoring each VR trial, providing real-time guidance and instructions to both patients. The introduction of the VR task to patients by their therapist, coupled with detailed explanations of its objectives and instructions, aligns with the literature on patient-centered education and rehabilitation practice [24]. The aspect of providing the patients with an understanding of the rehabilitation process may have contributed to both patients' positive feedback.

Individual Differences in Task Response

While it is important to keep in mind the anecdotal nature of the evidence due to this being a case study, tentatively positive results were obtained from the patient's interaction with the VR task. The integration of task metrics, standardized motor assessments, and patient feedback highlights individual differences in response to the VR intervention. Patient B showed a significant decay trend in task completion times, along with reduced intrasession variability and subjective reports of improved confidence in walking [17,86]. Patient A, by contrast, exhibited variable completion times and no perceived functional change. Pre- and postintervention scores on the BBT and 9HPT revealed mild left-hand impairment for both patients, with little measurable change over the 12 sessions—suggesting that motor gains alone are unlikely to account for Patient B's improved task efficiency. Notably, both patients performed within normative ranges on the 9HPT [87] for the unaffected hand, with lower scores for the affected hand, consistent with moderate upper-limb asymmetry typical of right hemisphere stroke. Furthermore, across sessions, ball bounce speed (see Table S3 in Multimedia Appendix 1, a proxy for task difficulty) was held constant, indicating that performance differences were not attributable to variation in task demands.

Although the distinct patient profiles, including comorbidities such as VSN, left-side hemiparesis, and left-side hemiplegia, may have influenced task interactions and completion times, it is critical to approach these findings with caution given the case study design. VSN involves attention and awareness deficits with perceptual components [6-8,88], while left-side hemiparesis and hemiplegia relate primarily to motor capacity, with hemiparesis indicating muscle weakness and hemiplegia signifying a complete loss of motor control. In addition, differences in the timing of intervention—Patient A, 1 year poststroke, versus Patient B, 4 years poststroke—may have contributed to variations in task outcomes. Therefore, it is essential to interpret these results as case-specific and within the limitations inherent to a case study framework.

The difference in the total VR task completion times, measured in days for the 2 patients, Patient A (74 days from Day 1 to the final session) and Patient B (42 days over the same period), tentatively suggests a shorter interval between treatment sessions is associated with faster task completion time. This finding is generally supported by the literature [21,22].

Limitations and Future Research

The study has several significant limitations that must be considered case-specific when interpreting its results. A key constraint for the quantitative data is that these are case studies of 2 patients diagnosed with VSN. This limited participant pool, as well as limited characterization of VSN symptoms, is limited by the absence of standardized neglect assessments (eg, the Behavioral Inattention Test) and neuroimaging data (eg, magnetic resonance imaging/computed tomography lesion localization), constraining the interpretation of the underlying neural correlates of task performance in response to this VR task. Furthermore, the 2 patients were at different stages in their rehabilitation journey—one having experienced a stroke in 2018 and the other in 2021. Such disparities in their recovery timelines could introduce confounding variables, thereby impacting the generalizability of the study's findings. In addition, the task-specific outcomes for each patient might have been influenced by numerous uncontrolled variables within the study.

To address these limitations, our future research roadmap involves research including a full sample of patients with VSN (a sample size calculation will be conducted using G*Power software, Heinrich-Heine-Universität Düsseldorf [89], derived from comparable VR rehabilitation studies), as well as additional standardized assessment measures (eg, the Montreal Cognitive Assessment [90], Hospital Anxiety and Depression Scale [91], and Berg Balance Scale [92] will be used). In addition, we will incorporate neglect-specific assessments, such as the Catherine Bergego Scale [93] and Behavioral Inattention Test [94] to quantify changes in VSN. To further investigate the impact of audiovisual cueing, we also plan to incorporate a comparison condition, such as a visual-only cueing group, alongside a standard control group. This will allow us to more rigorously evaluate the specific contribution of multisensory feedback to rehabilitation outcomes.

Regarding the VR task, future analyses of task completion time at varying locations across the 30° horizontal plane will be made to compare how patient progress in reaching easier versus harder targets on the contralesional side. A subsequent study may be designed to include trials on both the left and right sides, incorporating a structured analysis to determine whether the observed improvements in task performance are due to overall speed enhancement or are specifically observed within the neglected hemispace. Although out of scope of this study, collecting richer qualitative reports from family members who regularly interact with each patient will provide clearer insights into the patients' daily activities and recovery needs [3,95], offering a holistic perspective on their progress. While this study did not focus on improvements in grasping skill, methodologies from related research [28,31] could be applied in future studies to assess such skill in patients with VSN using this VR task. In addition, using the Suite for the Assessment of Low-Level Cues on Orientation [96] could enrich our understanding of how patients with VSN perceive and navigate in VR environments, potentially offering valuable insights for rehabilitation practices.

A follow-up study will include eye-tracking measurements to track visual attention during a trial, as well as include targeted evaluations of audiovisual cue clarity and perceived utility to better understand their functional role in attention orientation and motor engagement (eg, prompting gaze shifts or initiating reach movements). These improvements in study design and data collection will mitigate some of the limitations observed in this study.

Although participants did not report any discomfort or restrictions associated with the HMD equipment, future studies could explore the use of wireless HMD systems, such as the HTC Vive XR Elite, Meta Quest, or Varjo VR, to enhance patient comfort and mobility. In addition, incorporating advanced tracking technology in future studies could improve control over experimental variables.

An additional consideration for future exploration is the potential of the VR intervention to positively influence patient motivation. Both patients in our case studies responded very positively to the VR intervention, and for follow-up studies, measures of patient motivation such as the Motivation in Stroke Patients for Rehabilitation Scale [97] could be included to provide valuable insight into this aspect of the intervention. Crucially, to strengthen the evaluation of patient experience, future studies will incorporate validated instruments, such as the System Usability Scale [67] and Intrinsic Motivation Inventory [98], alongside direct patient-reported outcome measures, to reduce reliance on therapist-transcribed responses and minimize potential reporting bias. Motivation and patient enjoyment have been shown to be important drivers of positive rehabilitation outcomes [5]. The gamified nature of the VR task, coupled with the novelty of the technology, may support patient motivation, encouraging patients to persist in therapy despite the difficulty of the task.

Conclusions

This study explored the development and implementation of a VR-based physiotherapy intervention designed for hand grasping training for VSN rehabilitation. Positive preliminary user experience reports from both patients and physiotherapists provide promising evidence for a future research roadmap of this VR task and highlight the individual patient differences in response to VR-assisted physiotherapy. In addition, the distinct responses of the 2 patients highlight the intervention's potential capacity for personalized adaptation, emphasizing its suitability for diverse VSN rehabilitation needs. To further enhance the feasibility of this VR task in physiotherapy rehabilitation, future research should focus on the use of additional standardized measures applied to a full-scale sample. To better understand the specific impact of this task on attention to the neglected hemispace, future investigations should include a structured analysis of motor performance at varying target locations across the 30° horizontal plane. Furthermore, a full-scale sample could provide further insight into whether observed improvements are due to enhanced movement speed or specific engagement with the neglected hemispace. These findings support the use of VR as a patient-centered tool that can be tailored to individual profiles, offering promising directions for future research in neurorehabilitation.

Acknowledgments

This project has received funding from the Research Council of Finland (346210).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials on virtual reality setup and sessions.
[DOCX File, 494 KB - xr_v2i1e72439_app1.docx]

References

1. Esposito E, Shekhtman G, Chen P. Prevalence of spatial neglect post-stroke: a systematic review. Ann Phys Rehabil Med 2021 Sep;64(5):101459. [doi: 10.1016/j.rehab.2020.10.010] [Medline: 33246185]
2. Moore KS. A systematic review on the neural effects of music on emotion regulation: implications for music therapy practice. J Music Ther 2013;50(3):198-242. [doi: 10.1093/jmt/50.3.198] [Medline: 24568004]
3. Hesamzadeh A, Dalvandi A, Bagher Maddah S, Fallahi Khoshknab M, Ahmadi F, Mosavi Arfa N. Family caregivers' experience of activities of daily living handling in older adult with stroke: a qualitative research in the Iranian context. Scand J Caring Sci 2017 Sep;31(3):515-526. [doi: 10.1111/scs.12365] [Medline: 27530936]
4. Mubin O, Alnajjar F, Al Mahmud A, Jishtu N, Alsinglawi B. Exploring serious games for stroke rehabilitation: a scoping review. Disabil Rehabil Assist Technol 2022 Feb;17(2):159-165. [doi: 10.1080/17483107.2020.1768309] [Medline: 32508187]
5. Luque-Moreno C, Kiper P, Solís-Marcos I, et al. Virtual reality and physiotherapy in post-stroke functional re-education of the lower extremity: a controlled clinical trial on a new approach. J Pers Med 2021 Nov 16;11(11):1210-1210. [doi: 10.3390/jpm11111210] [Medline: 34834562]

6. Parton A, Malhotra P, Husain M. Hemispatial neglect. J Neurol Neurosurg Psychiatry 2004 Jan;75(1):13-21. [Medline: 14707298]
7. Driver J, Vuilleumier P. Perceptual awareness and its loss in unilateral neglect and extinction. Cognition 2001 Apr;79(1-2):39-88. [doi: 10.1016/s0010-0277(00)00124-4] [Medline: 11164023]
8. Pavani F, Husain M, Ládavas E, Driver J. Auditory deficits in visuospatial neglect patients. Cortex 2004 Apr;40(2):347-365. [doi: 10.1016/s0010-9452(08)70130-8] [Medline: 15156793]
9. Chechlacz M, Rotshtein P, Humphreys GW. Neuroanatomical dissections of unilateral visual neglect symptoms: ALE meta - analysis of lesion - symptom mapping. Front Hum Neurosci 2012;6:230. [doi: 10.3389/fnhum.2012.00230] [Medline: 22907997]
10. Husain M, Kennard C. Visual neglect associated with frontal lobe infarction. J Neurol 1996 Sep;243(9):652-657. [doi: 10.1007/BF00878662] [Medline: 8892067]
11. Bird CM, Malhotra P, Parton A, Coulthard E, Rushworth MFS, Husain M. Visual neglect after right posterior cerebral artery infarction. J Neurol Neurosurg Psychiatry 2006 Sep;77(9):1008-1012. [doi: 10.1136/jnnp.2006.094417] [Medline: 16772354]
12. Hildebrandt H, Spang K, Ebke M. Visuospatial hemi-inattention following cerebellar/brain stem bleeding. Neurocase 2002;8(4):323-329. [doi: 10.1076/neur.8.3.323.16197] [Medline: 12221145]
13. Kim J, Nam Y, Lee J, Suh YJ, Hwang I. ProxiFit: proximity magnetic sensing using a single commodity mobile toward holistic weight exercise monitoring. In: Proc ACM Interact Mob Wearable Ubiquitous Technol: Association for Computing Machinery; 2023, Vol. 7. [doi: 10.1145/3610920]
14. Silveri MC, Misciagna S, Terrezza G. Right side neglect in right cerebellar lesion. J Neurol Neurosurg Psychiatry 2001 Jul;71(1):114-117. [doi: 10.1136/jnnp.71.1.114] [Medline: 11413276]
15. Karnath HO, Himmelbach M, Rorden C. The subcortical anatomy of human spatial neglect: putamen, caudate nucleus and pulvinar. Brain (Bacau) 2002 Feb;125(Pt 2):350-360. [doi: 10.1093/brain/awf032] [Medline: 11844735]
16. Cogné M, Guillaud E, Guillot L, et al. Association between prism adaptation and auditory cues on spatial navigation in individuals with unilateral neglect. Ann Phys Rehabil Med 2020 Jan;63(1):12-20. [doi: 10.1016/j.rehab.2019.03.005] [Medline: 31009802]
17. Strelnikov K, Rosito M, Barone P. Effect of audiovisual training on monaural spatial hearing in horizontal plane. PLoS ONE 2011 Mar 29;6(3):e18344. [doi: 10.1371/journal.pone.0018344] [Medline: 21479241]
18. Elor A, Kurniawan S, Teodorescu M. Towards an immersive virtual reality game for smarter post-stroke rehabilitation. Presented at: 2018 IEEE International Conference on Smart Computing (SMARTCOMP); Jun 18-20, 2018; Taormina. [doi: 10.1109/SMARTCOMP.2018.00094]
19. Fransen M. When is physiotherapy appropriate? Best Pract Res Clin Rheumatol 2004 Aug;18(4):477-489. [doi: 10.1016/j.berh.2004.03.006] [Medline: 15301982]
20. Kempitiya T, De Silva D, Rio E, Skarbez R, Alahakoon D. Personalised physiotherapy rehabilitation using artificial intelligence and virtual reality gaming. Presented at: 2022 15th International Conference on Human System Interaction (HSI); Jul 28-31, 2022; Melbourne, Australia. [doi: 10.1109/HSI55341.2022.9869454]
21. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008 Feb;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]
22. Lang CE, Macdonald JR, Reisman DS, et al. Observation of amounts of movement practice provided during stroke rehabilitation. Arch Phys Med Rehabil 2009 Oct;90(10):1692-1698. [doi: 10.1016/j.apmr.2009.04.005] [Medline: 19801058]
23. Ekman I, Swedberg K, Taft C, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011 Dec;10(4):248-251. [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
24. Wijma AJ, Bletterman AN, Clark JR, et al. Patient-centeredness in physiotherapy: What does it entail? A systematic review of qualitative studies. Physiother Theory Pract 2017 Nov 2;33(11):825-840. [doi: 10.1080/09593985.2017.1357151]
25. Bollea L, Rosa GD, Gisondi A, et al. Recovery from hemiparesis and unilateral spatial neglect after neonatal stroke. Case report and rehabilitation of an infant. Brain Inj 2007 Jan;21(1):81-91. [doi: 10.1080/02699050601148882] [Medline: 17364523]
26. Elshout JA, Nijboer TCW, Van der Stigchel S. Is congruent movement training more effective than standard visual scanning therapy to ameliorate symptoms of visuospatial neglect? Study protocol of a randomised control trial. BMJ Open 2019 Dec 4;9(12):e031884. [doi: 10.1136/bmjopen-2019-031884] [Medline: 31806612]
27. van Wyk A, Eksteen CA, Rheeder P. The effect of visual scanning exercises integrated into physiotherapy in patients with unilateral spatial neglect poststroke: a matched-pair randomized control trial. Neurorehabil Neural Repair 2014;28(9):856-873. [doi: 10.1177/1545968314526306] [Medline: 24633138]
28. Dabholkar T, Shah SS. Effect of virtual reality training using a Leap Motion Controller on impairments and disability in patients with wrist and hand stiffness. Indian J Public Health Res Dev 2020;11(11):210-219. [doi: 10.37506/ijphrd.v11i11.11375]
29. Lorussi F, Carbonaro N, De Rossi D, Paradiso R, Veltink P, Tognetti A. Wearable textile platform for assessing stroke patient treatment in daily life conditions. Front Bioeng Biotechnol 2016;4:28. [doi: 10.3389/fbioe.2016.00028] [Medline: 27047939]

30. Rossit S, Fraser JA, Teasell R, Malhotra PA, Goodale MA. Impaired delayed but preserved immediate grasping in a neglect patient with parieto-occipital lesions. Neuropsychologia 2011 Jul;49(9):2498-2504. [doi: 10.1016/j.neuropsychologia.2011.04.030] [Medline: 21569783]
31. Turton AJ, Cunningham P, van Wijck F, et al. Home-based Reach-to-Grasp training for people after stroke is feasible: a pilot randomised controlled trial. Clin Rehabil 2017 Jul;31(7):891-903. [doi: 10.1177/0269215516661751] [Medline: 27470470]
32. Sisto SA, Forrest GF, Glendinning D. Virtual reality applications for motor rehabilitation after stroke. Top Stroke Rehabil 2002;8(4):11-23. [doi: 10.1310/YABD-14KA-159P-MN6F] [Medline: 14523727]
33. Pintoa S, Nicolau H, Lopes DS, Pereira AM. Interactive technologies in stroke recovery: uncovering challenges and opportunities through physiotherapist's perspective. Ann Med 2021 Apr;53(sup1):S3. [doi: 10.1080/07853890.2021.1893582]
34. Jang SH, You SH, Hallett M, et al. Cortical reorganization and associated functional motor recovery after virtual reality in patients with chronic stroke: an experimenter-blind preliminary study. Arch Phys Med Rehabil 2005 Nov;86(11):2218-2223. [doi: 10.1016/j.apmr.2005.04.015] [Medline: 16271575]
35. Schwab PJ, Miller A, Raphail AM, et al. Virtual reality tools for assessing unilateral spatial neglect: a novel opportunity for data collection. J Vis Exp 2021 Mar 10(169):e61951. [doi: 10.3791/61951] [Medline: 33779608]
36. Wagner S, Preim B, Saalfeld P, Belger J. Crossing ivroad: a VR application for detecting unilateral visuospatial neglect in poststroke patients. Presented at: 2019 International Conference on Virtual Rehabilitation (ICVR); Jul 21-24, 2019; Tel Aviv, Israel. [doi: 10.1109/ICVR46560.2019.8994643]
37. Moon HS, Shin SW, Chung ST, Kim E. K-CBS-based unilateral spatial neglect rehabilitation training contents utilizing virtual reality. Presented at: 2019 International Conference on Electronics, Information, and Communication (ICEIC); Jan 22-25, 2019; Auckland, New Zealand. [doi: 10.23919/ELINFOCOM.2019.8706395]
38. Aderinto N, Olatunji G, Abdulbasit MO, Edun M, Aboderin G, Egbunu E. Exploring the efficacy of virtual reality-based rehabilitation in stroke: a narrative review of current evidence. Ann Med 2023;55(2):2285907. [doi: 10.1080/07853890.2023.2285907] [Medline: 38010358]
39. Ansuini C, Pierno AC, Lusher D, Castiello U. Virtual reality applications for the remapping of space in neglect patients. Restor Neurol Neurosci 2006;24(4-6):431-441. [doi: 10.3233/RNN-2006-00350] [Medline: 17119316]
40. Castiello U, Lusher D, Burton C, Glover S, Disler P. Improving left hemispatial neglect using virtual reality. Neurology (ECronicon) 2004 Jun 8;62(11):1958-1962. [doi: 10.1212/01.wnl.0000128183.63917.02] [Medline: 15184596]
41. Knobel SEJ, Kaufmann BC, Geiser N, et al. Effects of virtual reality-based multimodal audio-tactile cueing in patients with spatial attention deficits: pilot usability study. JMIR Serious Games 2022 May 25;10(2):e34884. [doi: 10.2196/34884] [Medline: 35612894]
42. Leitner MC, Hawelka S. Visual field improvement in neglect after virtual reality intervention: a single-case study. Neurocase 2021 Jun;27(3):308-318. [doi: 10.1080/13554794.2021.1951302] [Medline: 34278959]
43. Ramos AA, Hørning EC, Wilms IL. Simulated prism exposure in immersed virtual reality produces larger prismatic after-effects than standard prism exposure in healthy subjects. PLoS ONE 2019;14(5):e0217074. [doi: 10.1371/journal.pone.0217074] [Medline: 31125360]
44. Frassinetti F, Angeli V, Meneghello F, Avanzi S, Làdavas E. Long-lasting amelioration of visuospatial neglect by prism adaptation. BRAIN (Bacau) 2002 Mar;125(Pt 3):608-623. [doi: 10.1093/brain/awf056] [Medline: 11872617]
45. Posner MI. Measuring alertness. Ann N Y Acad Sci 2008;1129(1):193-199. [doi: 10.1196/annals.1417.011] [Medline: 18591480]
46. Finke K, Matthias E, Keller I, Müller HJ, Schneider WX, Bublak P. How does phasic alerting improve performance in patients with unilateral neglect? A systematic analysis of attentional processing capacity and spatial weighting mechanisms. Neuropsychologia 2012 May;50(6):1178-1189. [doi: 10.1016/j.neuropsychologia.2012.02.008] [Medline: 22386879]
47. Làdavas E, Tosatto L, Bertini C. Behavioural and functional changes in neglect after multisensory stimulation. Neuropsychol Rehabil 2022 May 28;32(5):662-689. [doi: 10.1080/09602011.2020.1786411]
48. Cooper N, Millela F, Cant I, White MD, Meyer G. Transfer of training-Virtual reality training with augmented multisensory cues improves user experience during training and task performance in the real world. PLoS ONE 2021;16(3):e0248225. [doi: 10.1371/journal.pone.0248225] [Medline: 33760859]
49. Salatino A, Zavattaro C, Gammeri R, et al. Virtual reality rehabilitation for unilateral spatial neglect: a systematic review of immersive, semi-immersive and non-immersive techniques. Neurosci Biobehav Rev 2023 Sep;152:105248. [doi: 10.1016/j.neubiorev.2023.105248] [Medline: 37247829]
50. Rengachary J, He BJ, Shulman GL, Corbetta M. A behavioral analysis of spatial neglect and its recovery after stroke. Front Hum Neurosci 2011;5:29. [doi: 10.3389/fnhum.2011.00029] [Medline: 21519374]
51. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
52. Danso A, Leandertz M, Ala-Ruona E, Rousi R. Neglect, virtual reality and music therapy: a narrative review. MMD 2022 Jul;14(3). [doi: 10.47513/mmd.v14i3.865]
53. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29:a1655. [doi: 10.1136/bmj.a1655]

54. Lafitte R, Jeager M, Piscicelli C, et al. Spatial neglect encompasses impaired verticality representation after right hemisphere stroke. Ann N Y Acad Sci 2023 Feb;1520(1):140-152. [doi: 10.1111/nyas.14938] [Medline: 36478572]
55. Golay L, Hauert CA, Greber C, Schnider A, Ptak R. Dynamic modulation of visual detection by auditory cues in spatial neglect. Neuropsychologia 2005 Jan;43(9):1258-1265. [doi: 10.1016/j.neuropsychologia.2004.12.006]
56. Mattingley JB, Husain M, Rorden C, Kennard C, Driver J. Motor role of human inferior parietal lobe revealed in unilateral neglect patients. Nature New Biol 1998 Mar;392(6672):179-182. [doi: 10.1038/32413]
57. Stammler B, Flammer K, Schuster T, Lambert M, Karnath HO. Negami: an augmented reality app for the treatment of spatial neglect after stroke. JMIR Serious Games 2023 Feb 27;11:e40651. [doi: 10.2196/40651] [Medline: 36848215]
58. Yoshizawa M, Yoshida Y, Baheux K, Tanaka A, Seki K, Handa Y. Development of virtual reality systems for tests and rehabilitation of patients with hemispatial neglect. Presented at: 2007 IEEE/ICME International Conference on Complex Medical Engineering; May 23-27, 2007; Beijing, china. [doi: 10.1109/ICCME.2007.4381956]
59. Dozio N, Maggioni E, Pittera D, Gallace A, Obrist M. May I smell your attention: exploration of smell and sound for visuospatial attention in virtual reality. Front Psychol 2021;12:671470. [doi: 10.3389/fpsyg.2021.671470] [Medline: 34366990]
60. Fernandes AS, Feiner SK. Combating VR sickness through subtle dynamic field-of-view modification. Presented at: 2016 IEEE Symposium on 3D User Interfaces (3DUI); Mar 19-20, 2016; Greenville, SC, USA. [doi: 10.1109/3DUI.2016.7460053]
61. Levin MF, Kleim JA, Wolf SL. What do motor “recovery” and “compensation” mean in patients following stroke? Neurorehabil Neural Repair 2009 May;23(4):313-319. [doi: 10.1177/1545968308328727] [Medline: 19118128]
62. Wilson FC, Manly T, Coyle D, Robertson IH. The effect of contralesional limb activation training and sustained attention training for self-care programmes in unilateral spatial neglect. Restor Neurol Neurosci IOS Press 2000;16(1):1-4. [doi: 10.3233/RNN-2000-00128]
63. MacKenzie IS. Fitts’ law. In: Norman KL, Kirakowski J, editors. The Wiley Handbook of Human Computer Interaction: Wiley; 2018:347-370. [doi: 10.1002/9781118976005]
64. Serra S, Kachach R, Gonzalez-Sosa E, Villegas A. Natural User Interfaces for Mixed Reality: Controlling Virtual Objects with Your Real Hands: Institute of Electrical and Electronics Engineers Inc; 2020:713-714. [doi: 10.1109/VRW50115.2020.00207]
65. Cameirão MS, Badia SI, Duarte E, Frisoli A, Verschure P. The combined impact of virtual reality neurorehabilitation and its interfaces on upper extremity functional recovery in patients with chronic stroke. Stroke 2012 Oct;43(10):2720-2728. [doi: 10.1161/STROKEAHA.112.653196] [Medline: 22871683]
66. Wang Y, Wu Y, Jung S, Hoermann S, Yao S, Lindeman RW. Enlarging the usable hand tracking area by using multiple Leap Motion Controllers in VR. IEEE Sensors J 2021 Aug 15;21(16):17947-17961. [doi: 10.1109/JSEN.2021.3082988]
67. Cheah WH, Mat Jusoh N, Aung MMT, Ab Ghani A, Mohd Amin Rebuan H. Mobile technology in medicine: development and validation of an adapted System Usability Scale (SUS) questionnaire and modified Technology Acceptance Model (TAM) to evaluate user experience and acceptability of a mobile application in MRI safety screening. Indian J Radiol Imaging 2023 Jan;33(1):36-45. [doi: 10.1055/s-0042-1758198] [Medline: 36855734]
68. Steinmaurer A, Sackl M, Gutl C. Engagement in in-game questionnaires - perspectives from users and experts. Presented at: 2021 7th International Conference of the Immersive Learning Research Network (iLRN); May 17 to Jun 10, 2021; Eureka, CA, USA. [doi: 10.23919/iLRN52045.2021.9459373]
69. Sexton JB, Helmreich RL, Neilands TB, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. BMC Health Serv Res 2006 Apr 3;6:44. [doi: 10.1186/1472-6963-6-44] [Medline: 16584553]
70. WMA Declaration of Helsinki – ethical principles for medical research involving human participants. World Medical Association. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki/ [accessed 2025-01-07]
71. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11(11):CD008349. [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
72. Paparini S, Green J, Papoutsi C, et al. Case study research for better evaluations of complex interventions: rationale and challenges. BMC Med 2020 Nov 10;18(1):301. [doi: 10.1186/s12916-020-01777-6] [Medline: 33167974]
73. Lloréns R, Gil-Gómez JA, Alcañiz M, Colomer C, Noé E. Improvement in balance using a virtual reality-based stepping exercise: a randomized controlled trial involving individuals with chronic stroke. Clin Rehabil 2015 Mar;29(3):261-268. [doi: 10.1177/0269215514543333] [Medline: 25056999]
74. Pazzaglia C, Imbimbo I, Tranchita E, et al. Comparison of virtual reality rehabilitation and conventional rehabilitation in Parkinson’s disease: a randomised controlled trial. Physiotherapy 2020 Mar;106:36-42. [doi: 10.1016/j.physio.2019.12.007] [Medline: 32026844]
75. Desrosiers J, Bravo G, Hébert R, Dutil E, Mercier L. Validation of the Box and Block Test as a measure of dexterity of elderly people: reliability, validity, and norms studies. Arch Phys Med Rehabil 1994 Jul;75(7):751-755. [doi: 10.1016/0003-9993(94)90130-9] [Medline: 8024419]
76. Robinson A. Applied asymptotics: case studies in small-sample statistics. J R Stat Soc Ser A Stat Soc 2008 Apr 1;171(2):504-505. [doi: 10.1111/j.1467-985X.2007.00528_3.x]
77. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]

78. Ng VKY, Cribbie RA. Using the gamma generalized linear model for modeling continuous, skewed and heteroscedastic outcomes in psychology. Curr Psychol 2017 Jun;36(2):225-235. [doi: 10.1007/s12144-015-9404-0]
79. Pua YH, Seah FJT, Poon CLL, et al. Age- and sex-based recovery curves to track functional outcomes in older adults with total knee arthroplasty. Age Ageing 2018 Jan 1;47(1):144-148. [doi: 10.1093/ageing/afx148] [Medline: 28985252]
80. Lüdecke D, Ben-Shachar M, Patil I, Waggoner P, Makowski D. performance: an R package for assessment, comparison and testing of statistical models. J Open Source Softw 2021;6(60):3139. [doi: 10.21105/joss.03139]
81. Rodgers JL. The epistemology of mathematical and statistical modeling: a quiet methodological revolution. Am Psychol 2010 Jan;65(1):1-12. [doi: 10.1037/a0018326] [Medline: 20063905]
82. Zhang J, Yang Y, Ding J. Information criteria for model selection. WIREs Computational Stats 2023 Sep;15(5):e1607. [doi: 10.1002/wics.1607]
83. Heyse J, Carlier S, Verhelst E, Vander Linden C, De Backere F, De Turck F. From patient to musician: a multi-sensory virtual reality rehabilitation tool for spatial neglect. Appl Sci (Basel) 2022 Jan;12(3):1242. [doi: 10.3390/app12031242]
84. Koo CH, Park JW, Ryu JH, Han SH. The effect of virtual reality on preoperative anxiety: a meta-analysis of randomized controlled trials. J Clin Med 2020 Sep 29;9(10):3151. [doi: 10.3390/jcm9103151] [Medline: 33003411]
85. Melander Wikman A, Fältholm Y. Patient empowerment in rehabilitation: “Somebody told me to get rehabilitated”. Adv Physiother 2006 Jan;8(1):23-32. [doi: 10.1080/14038190500494774]
86. Wåhlin A, Fordell H, Ekman U, Lenfeldt N, Malm J. Rehabilitation in chronic spatial neglect strengthens resting-state connectivity. Acta Neurol Scand 2019 Mar;139(3):254-259. [doi: 10.1111/ane.13048] [Medline: 30427058]
87. Mathiowetz V, Weber K, Kashman N, Volland G. Adult norms for the Nine Hole Peg Test of finger dexterity. Occup Ther J Res 1985 Jan;5(1):24-38. [doi: 10.1177/153944928500500102]
88. Moore MJ, Milosevich E, Mattingley JB, Demeyere N. The neuroanatomy of visuospatial neglect: a systematic review and analysis of lesion-mapping methodology. Neuropsychologia 2023 Feb 10;180:108470. [doi: 10.1016/j.neuropsychologia.2023.108470] [Medline: 36621594]
89. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
90. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J American Geriatrics Society 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x]
91. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
92. Katherine B. Measuring balance in the elderly: development and validation of an instrument [Thesis]. : McGill University; 1994.
93. Chen P, Hreha K, Fortis P, Goedert KM, Barrett AM. Functional assessment of spatial neglect: a review of the Catherine Bergego scale and an introduction of the Kessler foundation neglect assessment process. Top Stroke Rehabil 2012;19(5):423-435. [doi: 10.1310/tsr1905-423] [Medline: 22982830]
94. Hartman-Maeir A, Katz N. Validity of the Behavioral Inattention Test (BIT): relationships with functional tasks. Am J Occup Ther 1995 Jun;49(6):507-516. [doi: 10.5014/ajot.49.6.507] [Medline: 7645663]
95. Mayo NE, Wood-Dauphinee S, Côté R, Durcan L, Carlton J. Activity, participation, and quality of life 6 months poststroke. Arch Phys Med Rehabil 2002 Aug;83(8):1035-1042. [doi: 10.1053/apmr.2002.33984] [Medline: 12161823]
96. Esposito D, Bollini A, Gori M. The Suite for the Assessment of Low-Level cues on Orientation (SALLO): The psychophysics of spatial orientation in virtual reality. Behav Res Methods 2024 Aug;56(5):5214-5231. [doi: 10.3758/s13428-023-02265-4] [Medline: 37932625]
97. Yoshida T, Otaka Y, Osu R, Kumagai M, Kitamura S, Yaeda J. Motivation for rehabilitation in patients with subacute stroke: a qualitative study. Front Rehabilit Sci 2021;2. [doi: 10.3389/fresc.2021.664758]
98. Ryan RM, Mims V, Koestner R. Relation of reward contingency and interpersonal context to intrinsic motivation: a review and test using cognitive evaluation theory. J Pers Soc Psychol 1983;45(4):736-750. [doi: 10.1037//0022-3514.45.4.736]

Abbreviations

9HPT: 9-Hole Peg Test
BBT: Box and Block Test
FOV: field of view
HMD: head-mounted display
HRTF: head-related transfer function
MRC: Medical Research Council
VR: virtual reality
VSN: visuospatial neglect

Edited by L Riedemann; submitted 10.02.25; peer-reviewed by A Adeoye, G Heng, H Wang; revised version received 09.06.25; accepted 11.07.25; published 15.08.25.

Please cite as:
Danso A, Nijhuis P, Ansani A, Hartmann M, Minkkinen G, Luck G, Bamford JS, Faber S, Agres KR, Glasser S, Särkämö T, Rousi R, Thompson MR
Development and User Experiences of a Novel Virtual Reality Task for Poststroke Visuospatial Neglect: Exploratory Case Study
JMIR XR Spatial Comput 2025;2:e72439
URL: https://xr.jmir.org/2025/1/e72439
doi:10.2196/72439

A Local Training Program to Increase Awareness of Emerging Extended Reality Technologies Among Health Care Professionals: Development Study

Charlotte Galvin, MbChB, MA, MClinED; Jonathan Watt, BA; Payal Ghatnekar, BBA, MBA, MSc, PhD; Nicholas Peres*, BA, PhD; Jacqueline Rees-Lee*, MA, MPhil, MEd

Torbay and South Devon NHS Foundation Trust, Torbay Hospital, Torquay, United Kingdom

*these authors contributed equally

Corresponding Author:
Jacqueline Rees-Lee, MA, MPhil, MEd
Torbay and South Devon NHS Foundation Trust, Torbay Hospital, Torquay, United Kingdom

Abstract

Background: Demands on health care services can greatly outweigh capacity. Multifactorial causative factors present great challenges, forcing the National Health Service (NHS) to increase efficiency and adaptivity. Concurrently, digital advancements are excelling and long-term plans for NHS sustainability are focusing on the use of technological interventions to benefit patients. As a result, integration of extended reality (XR) technology has become an important focus of health care research. However, models of how the digital literacy of health care workforces can be developed and how frontline staff can be actively involved in the design and development of creative digital interventions are lacking. Such programs are essential to allow the development and upscaling of digital innovation within the NHS for the benefit of the patients. Such a program has been developed in the Digital Futures research lab at Torbay and South Devon NHS Foundation Trust, representing one of the first immersive digital technologies research spaces embedded within the NHS. A "Digital Deep Dive" training program has been developed, allowing local health care workers to recognize the possibilities of digital health care technologies and supporting them in the evolution of ideas for potential bespoke digital solutions appropriate to their own patient groups and care pathways.

Objective: This paper aims to explain the development of this unique XR Deep Dive program and present the evaluation that informed future directions for its ongoing development.

Methods: The Deep Dive sessions were designed according to relevant pedagogic principles, including experiential, active, and contextual learning theories. Voluntary pilot sessions were held for local clinical teams comprised of junior doctors, consultants, nurses, and allied health professionals. Self-selection sampling was used. Participants completed an anonymous postsession feedback form, which was used to conduct a service evaluation. Data were analyzed using descriptive statistics (quantitative) and thematic analysis (qualitative).

Results: In total, 21 completed questionnaires were analyzed. Overall, the sessions were positively received: all participants reported increased awareness of the potential for digital health care innovation postsession and most found it useful and relevant to their clinical careers. Participants valued the sessions being grounded in a context relevant to local practice with opportunities to interact with the technology through the lens of use cases.

Conclusions: We have developed a unique training initiative providing contextually relevant XR technology awareness training for health care professionals locally. Despite the growing pace of digital health care innovation, we recognized a knowledge gap in our local workforce regarding the potential of XR technologies within health care. We responded by developing a training program grounded in the concept of digital co-creation—working with staff and service users to develop bespoke solutions integrated within patient pathways. The results from this paper will help to inform future directions for developing digital awareness training in our trust and have implications for wider NHS digital literacy training.

(JMIR XR Spatial Comput 2025;2:e57361) doi:10.2196/57361

KEYWORDS

health care XR; extended reality in health care; XR; virtual reality in health care; VR; digital awareness training; digital deep dive; digital literacy; emerging health care technology; digital future; extended reality; virtual reality

Introduction

We live in an era where demands on NHS services can outweigh capacity. This mismatch in capacity versus demand is increasing and causative factors are multifactorial, including an aging population, significant years of underfunding, a reducing and inadequate workforce, and the COVID-19 pandemic. To meet these challenges, health care services must become more adaptive and efficient, while maintaining a world-leading standard of patient and clinician experience, service quality, and clinical safety. It is also an era where technological and digital advancements are progressing at an unprecedented rate.

The 2019 government-commissioned Topol Review [1] made important recommendations to ensure the NHS becomes a world leader in digital technologies utilization for the benefit of patients, and the necessity to grow the digital literacy of the health care workforce was further accelerated by the COVID-19 pandemic [2]. In a more recent development, the 2023 NHS Long Term Workforce Plan [3] underscores the significance of digital competencies and integration as crucial components in equipping the workforce to meet prospective service demands.

Extended reality (XR)—an umbrella term encapsulating the spectrum of immersive technologies from simple augmented reality (AR) through to complete virtual reality (VR)—has become a key focus of cutting-edge health care research [4], with its benefits becoming clearer through use in as many as 97 UK health organizations and 119 distinct health care research projects in 2021 [5]. The comparison of XR-driven practices to traditional methods in medicine [6,7], surgery [8,9], rehabilitation [10,11], and clinical education [12,13] have become important research foci in recent years. Ultimately, the effectiveness of XR technologies in enhancing clinical skills and patient outcomes has been well demonstrated [6]. However, as important as these research projects are, they are insufficient if not accompanied by programs of digital training and education to reach the wider workforce.

A review of the literature has indicated that, while studies exploring the use of XR in a health care setting are numerous, real-world working models of health care workforce XR awareness training are lacking, with no applicable papers yielded from our search. Thus, despite the advancements in XR technologies within health care, there is a notable gap in the literature regarding the training of health care professionals to effectively integrate these tools into clinical practice for the benefit of patients. We propose that in order for XR technologies to be truly embedded in the NHS, within clinical care pathways and for the benefit of patients, they need to be understood and utilized by clinicians and health care professionals within the correct health care context. Although many digital technology companies are innovating in this space, direct access to and collaboration with clinicians and patients from the first stage of their innovation is lacking, meaning there is often a mismatch or lack of true co-design in what is being developed and what is actually required.

In 2020, this paper's senior authors (JRL and NP) were profoundly aware of the lack of digital literacy within their local NHS health care workforce and the lack of successful fully integrated digital-clinical partnerships. Working together and alongside other experts to allow a true understanding of both the clinical and digital worlds, they set out to address this by conceptualizing and developing the unique "Digital Futures: Human Centred Digital Innovation" program [14], which was initially supported by funding from Health Education England. The idea was to allow a true understanding of both the clinical and digital worlds and develop innovations in the common ground between their areas of expertise. Thus, the conceptualization of a Digital Futures Research Lab built on an existing XR Lab, which had been in development at Torbay Hospital since 2016. The program represents one of the first immersive digital technologies research spaces embedded within the NHS to inform national insights into research and development of immersive digital technologies in health care.

The development of a "Digital Deep Dive" training program was one of the founding principles of the Digital Futures program. Its aim is to increase digital literacy and awareness in local clinical teams, supporting them to recognize the possibilities of digital health care technologies and evolve ideas for potential bespoke digital solutions appropriate to their own patient groups. The clinical user-led approach of joining digital experts and clinical experts was conceptualized to allow cross-fertilization of ideas and knowledge to support the creation of bespoke solutions within the patient pathways and represents a "bottom up" approach of educating staff groups in digital technology, which is now gaining national interest.

Through this paper, we aim to highlight how we have developed local XR Deep Dive Training Sessions as part of the Digital Futures Programme and evaluate the impact of pilot sessions we have delivered.

Methods

Design

The XR Deep Dive training sessions have been developed collaboratively between clinicians and digital experts at Torbay and South Devon Foundation Trust (TSDFT). The sessions were designed to be delivered to teams of health care professionals across the trust in the on-site TSDFT Digital Futures Research Lab. Since the authors consider cross-fertilization of digital and clinical expertise to be paramount in the development of digital interventions that are useful and usable in practice, the sessions were designed to be co-delivered by a clinician and a digital expert.

The Deep Dive learning strategy was originally conceptualized by a global learning design company in the early 2000s and has since been widely implemented across various industries to promote learning and process development within professional teams [15]. Core to the Deep Dive methodology is integration of key stakeholders, affording them the opportunity to experiment with new concepts and brainstorm how that concept could be adapted and successfully integrated into their own unique context [15]. This approach offers an ideal solution to the challenges of XR health care training we have previously described. Therefore, we have adapted the Deep Dive methodology to develop our local training program: we first

introduce participants to the concept of XR, then we demonstrate its potential within health care, and finally we allow time, space, and support for teams to explore how the concept could be developed within the context of their own health care specialty for the direct benefit of local teams and patients.

To achieve this, we grounded our Digital Deep Dive session design in Experiential Learning theory [16]. A vital component of the deep dives is to showcase examples of embedded digital technologies in health care pathways across both our own trust and more widely, thus feeding the imaginations of the participants with the possibilities within the digital health care space by promoting hands-on experience and reflection [16]. In-session digital interaction was a key design priority, with time allocated to practical demonstrations and "digital playtime" allowing participants to trial the XR technology first hand. This also aligns with active learning theories and evidence that this type of digital interaction is a key component of achieving successful technology training [17]. The Digital Futures program has a "human first" approach to all its innovations, emphasizing how digital innovation can be utilized directly to improve patient care. In the Deep Dives, we therefore focus on technology in a humanistic sense—adopting this approach accentuates the personal, emotional, and psychological needs of the person in addition to their physical health needs, stressing the importance of treating each person as a unique individual, ensuring that care is patient-centred and that the health care experience is characterized by compassion, empathy, respect, and dignity [18]. We aimed to showcase how technology can be used to connect us with and value one another as fellow human beings, and so incorporated illustration of local use cases to provide context and authenticity. This design choice aligns with the goal of uniting concept with practice, which is central to contextual teaching and learning theory [19]. The informal learning environment was designed to encourage questions and discussion throughout, thereby supporting learners to develop a deeper understanding and explore different perspectives [20]. Time was also allotted at the end of the session for a mini focus group to further promote ideas for co-design and interdisciplinary collaboration of potential digital solutions. Sessions were designed to be delivered in a small group format (<10 participants), as this has been shown to foster better group collaboration, interaction, and discussion [20]. Finally, given the importance of posttraining follow-up to provide further support and ensure ongoing development [21], we considered how we would deliver postsession support as part of our program design—signposting to digital drop-in clinics to further improve targeted digital skills and share and refine ideas for future digital innovation was therefore promoted at the end of the Deep Dive sessions.

These design principles for the XR Deep Dive session are outlined in Figure 1, encapsulating the overarching aims of the training sessions, which are summarized in Figure 2.

Following the design phase, 8 voluntary pilot sessions were held between May 2022 and May 2023. Health care professionals—including resident doctors, consultants, nurses, occupational therapists, play specialists, and physiotherapists—from departments across TSDFT were invited between May 2022 and April 2023 via email and online trust advertising platforms to attend on a voluntary basis, therefore utilizing self-selection sampling [22]. Volunteers from all of these clinical groups attended sessions, with each session hosting between 3 and 8 participants to maintain the important small group sizes. Participants were invited to complete an anonymous postsession QR feedback form in Multimedia Appendix 1; by submitting this, participants consented for their anonymized comments to be included in this service evaluation. The Squire Guideline for Service Evaluation was used as a framework [23]. Free-text responses were evaluated by 2 authors (CG and PG) using thematic analysis, which is the accepted preferred method of interpreting qualitative data [24].

Each session was also observed by the senior author (JRL), who provided feedback on content and flow and suggested modifications. Using this feedback combined with the participant feedback, through an iterative process, the final content of the Digital Deep Dive sessions took shape.

Figure 1. An outline of the design of the XR deep dive training sessions. VR: virtual reality; XR: extended reality.

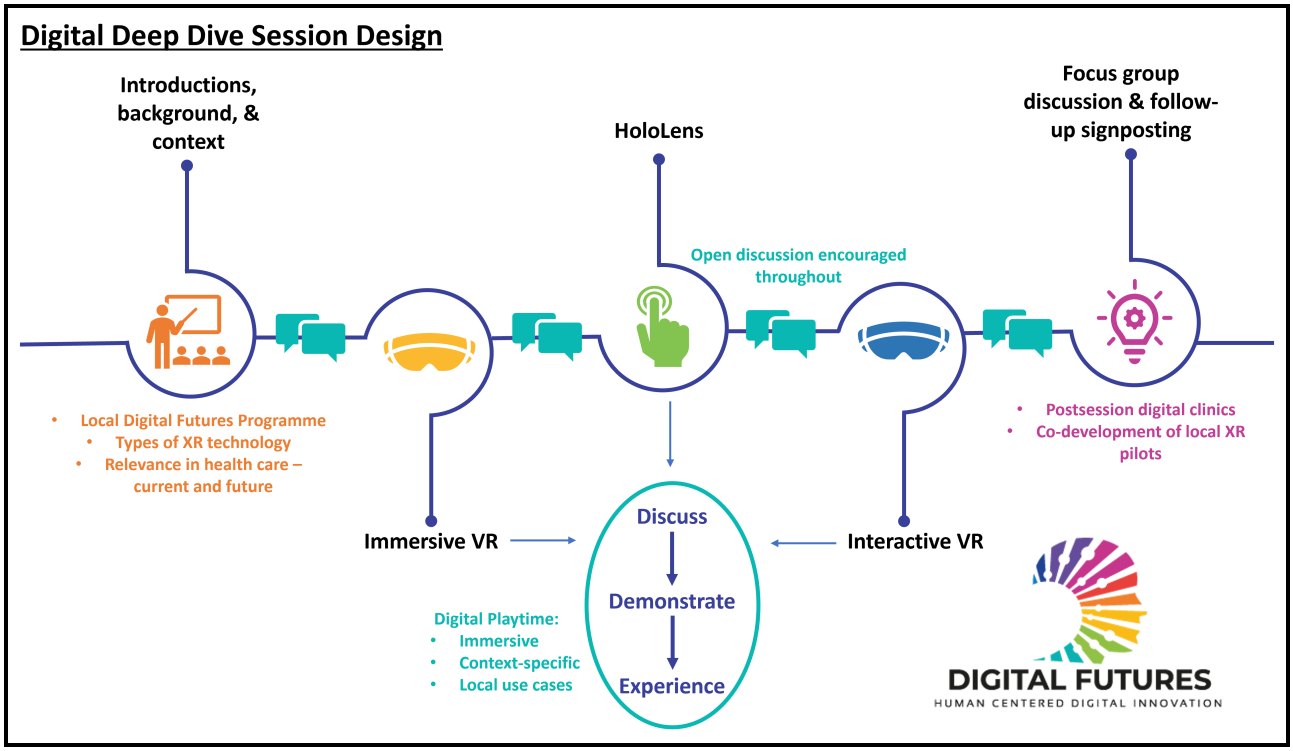
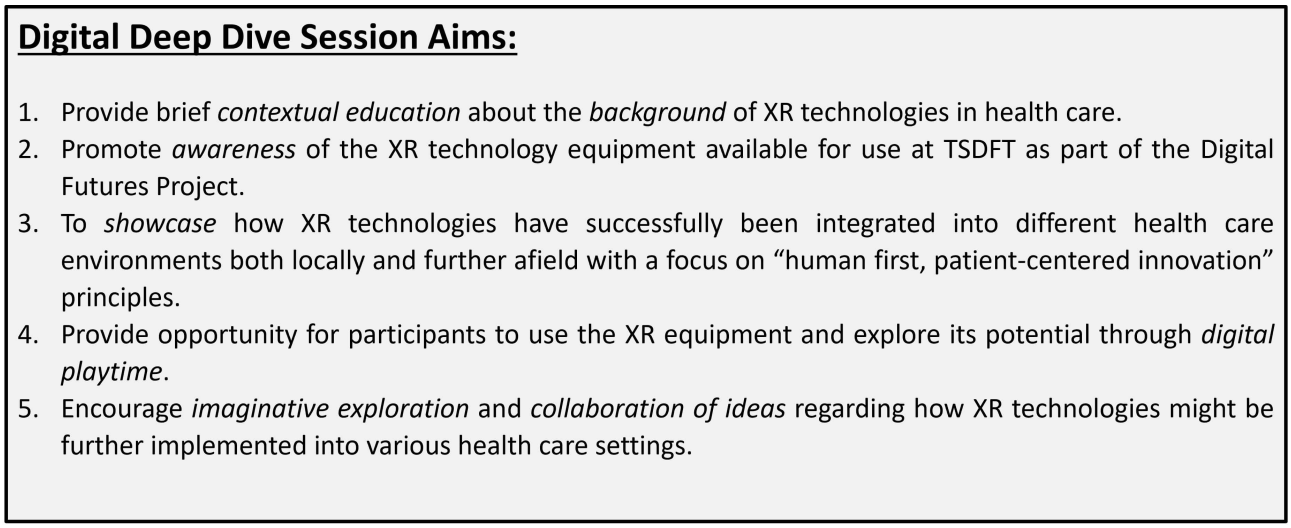


Figure 2. The aims of the XR deep dive training sessions. TSDFT: Torbay and South Devon Foundation Trust; XR: extended reality.



Ethical Considerations

In line with guidance provided by the Health Research Authority and compatible local Research and Development policies at TSDFT, a formal ethics application was not required for this service evaluation project. Participants were made aware through a formal statement on the feedback form that their anonymous responses may be used for evaluation purposes and may be included in future published work.

Results

From a total of 8 sessions delivered to 35 participants, 21 completed questionnaires were received, with a mix of qualitative and quantitative responses (60% response rate). Quantitative responses were analyzed using descriptive statistics and free-text responses were thematically grouped and analyzed.

Quantitative Data

Data were collected through a series of closed questions and 5-point Likert scales. Quantitative data were collected in 2 categories: presession experience and postsession feedback.

Presession Experience

Results are displayed in Table 1. All participants who took part in the XR Deep Dive sessions had little to no experience of using XR technology previously. Although just over half of participants were aware of XR being used in a health care context—either generally or specifically—the remainder had never heard of XR technologies being implemented in health care, and none had any personal involvement in using XR

technologies in a health care context. Further, most participants had never heard of the Digital Futures Programme at TSDFT and knew nothing or very little about current use of XR technologies in our local health care services.

Table . Quantitative data (presession ideas).

Question and answer	Number of responses (N=21)	Percentage of total responses
Before this session, what was your experience with virtual reality/augmented reality technologies?		
I had used these technologies a few times previously	11	52
I had heard of these technologies but had never used them	9	43
I had never heard of these technologies before	1	5
I had lots of experience of using these technologies	0	0
Before this session, how familiar were you with the use of digital technologies such as virtual reality/augmented reality in health care environments?		
I had never heard of these technologies being used in health care before	8	38
I had heard of these technologies being utilized in health care but did not have much knowledge regarding how	7	33
I had heard about specific projects involving these technologies in health care but have had no personal involvement	6	29
I have personally been involved in projects utilizing these technologies in health care settings	0	0
On a scale of 1 - 5, how much did you previously know about the digital projects ongoing at Torbay and South Devon Foundation Trust?		
1 (absolutely nothing)	16	76
2	2	10
3	3	14
4	0	0
5 (expert)	0	0
Had you previously heard of the Digital Futures Programme?		
Yes, and I knew what it was	1	4.76
Yes, but I didn't know what it was	1	4.76
No	19	90.48

Postsession Feedback

Results are displayed in Table 2. All participants indicated that they had a better understanding of the Digital Futures Programme and ongoing XR projects within the trust after taking part in the session. Most participants felt that the session was both useful and relevant to their future clinical careers and reported feeling inspired or very inspired to utilize XR technologies in their own health care specialty. Most participants indicated that they felt to some degree more confident in operating the XR equipment after the session.

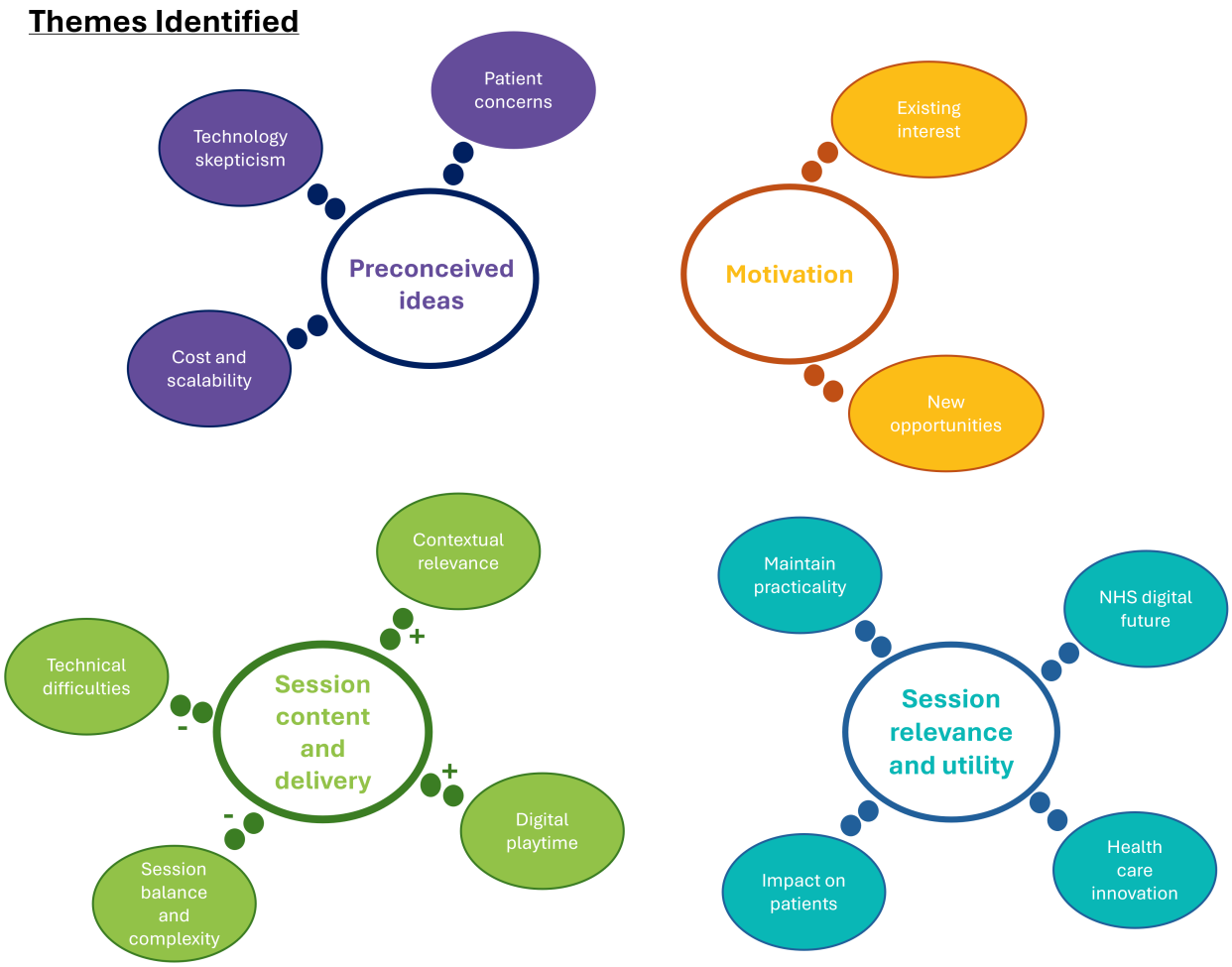
Table . Quantitative data (postsession feedback).

Question and answer	Number of responses (N=21)	Percentage of total responses
Do you now have a better understanding of the Digital Futures Programme and the current digital projects ongoing in Torbay?		
Yes	21	100
No	0	0
On a scale of 1‐5, do you feel this session has inspired some ideas for how you might utilize digital technology in your chosen health care specialty?		
1 (not at all)	0	0
2	0	0
3	2	10
4	7	33
5 (completely)	12	57
On a scale of 1‐5, how likely would you now be to get involved in a digital technologies in health care project in the future?		
1 (extremely unlikely)	0	0
2	0	0
3	4	19
4	7	33
5 (extremely likely)	10	48
On a scale of 1‐5, how much more confident do you now feel in operating the virtual reality/HoloLens technologies compared to before the session?		
1 (not any more confident)	0	0
2	0	0
3	5	24
4	13	62
5 (entirely more confident)	3	14
Do you think this session was useful to your future career?		
Yes	20	95
No	0	0
Unsure	1	5
Do you think this session was relevant to your future career?		
Yes	20	95
No	0	0
Unsure	1	5

Free-Text Data

Free-text responses were collected in 4 main areas: presession ideas and motivation, session content and delivery; session relevance and utility; and postsession development. Following thematic analysis of the responses, key themes were identified in each of these areas. These themes are presented visually in Figure 3.

Figure 3. A visual representation of the themes identified from the free-text responses.



Presession Ideas and Motivation

Participants were asked 2 free-text questions in this area—the first related to presession ideas about technology use in health care and the second related to why the participant chose to get involved in a Deep Dive session.

Of the 21 respondents, 13 (62%) raised preconceived ideas about use of XR technology in health care. From these responses, 3 themes were identified: Patient Concerns, Technology Skepticism, and Cost and Scalability.

First, responses from 8 participants included concerns that highlight the preconceived ideas that technology would damage the patient-clinician relationship; technology use would lead to impersonal health care; and technology would present usability issues in certain patient groups, such as older patients. Together, these answers contribute to the dominant theme of Patient Concerns. Presented below are some direct quotes from the participants:

> *I wondered how user-friendly the equipment might be, especially for older patients.* [Participant 4]
>
> *Worried about virtual technology replacing physical examination with patients.* [Participant 19]
>
> *Negative impact on the clinician-patient relationship—not very personal.* [Participant 10]

Second, Technology Skepticism emerged as another preconceived idea. Participants expressed valid concerns about the relative infancy of XR technologies, particularly XR for health care, with some participant responses presented below:

> *I know of such technology in the gaming world, but...I was skeptical about its uses in healthcare.* [Participant 7]
>
> *Technology and its use in healthcare are still very much in their infancy.* [Participant 5]

The third theme that emerged from asking about preconceived ideas is that of Cost and Scalability. Four participants raised the concern that digital projects in health care may be unrealistic due to the costs involved, and its impact on availability and accessibility to the technologies. Some of the responses from the survey participants are presented below:

> *Very costly so thought it would not be very achievable on a large scale.* [Participant 8]
>
> *Funding is likely to be the big barrier.* [Participant 7]

Next, the motivation of respondents to participate in the Deep Dive sessions fell into 2 themes: Exploring an Existing Interest and Curiosity About New Opportunities. In response to the question about motivation for participating in the sessions, words such as “exciting,” “interesting,” and “unique” were used frequently.

An existing interest in digital technology was identified by 7 participants as motivation for their involvement in the training sessions. One participant stated:

> *I am creative. I already know a bit about tech. I agree there is huge potential in using technology, specifically VR, to help people.* [Participant 2]

Further, 13 participants talked about being curious about what they perceived to be a new and interesting area. Multiple participants alluded to technology being part of the future in health care and that it holds many opportunities for development. Some quotes from the participants are presented below:

> *Interesting area of future development.* [Participant 20]
>
> *Wanted to hear more about what opportunity there was.* [Participant 12]

Session Content and Delivery

Participants were asked to identify the best thing about the session and whether they had any improvement suggestions. To ensure future session improvements, a specific question was also asked about any difficulties participants experienced when using the digital technology.

Positive comments about the session content and delivery were grouped in 2 themes: Digital Playtime and Contextual Relevance.

When asked to identify the best thing about the session, participants overwhelmingly gave answers that can be categorized into the theme of Digital Playtime. The hands-on digital experience integral to the session design was met with substantive positivity, with 19 of 21 participants (90%) citing the opportunity to use the technology in the session as one of the best aspects. Some example survey responses are below:

> *Fantastic to have hands on experience and understand more about how it all works.* [Participant 11]
>
> *Practical time with the headsets.* [Participant 13]

Next, participants particularly valued the use of local case studies to illustrate real-life application and contextual relevance, with 6 participants commenting that integration of use cases into the session was one of its best aspects. One participant said it was:

> *Brilliant to see the difference it's already making in the trust and the collaboration and partnership working already going on.* [Participant 11]

Participants were then asked about any specific technology difficulties experienced during the session and whether they had any improvement suggestions. Regarding technology difficulties, participants outlined 4 problems: connectivity issues (6 participants), motion sickness/nausea (2 participants), device fit issues (2 participants), and time to adjust (2 participants).

Eleven of 21 participants (52%) then made suggestions for session improvement. From the responses, 3 themes emerged: improvement of session balance, improvement of session complexity, and improvement of internet connectivity.

First, 8 participants gave answers that indicated better session balance would be welcomed. Integrating more digital playtime and less presentation time was frequently cited. Some participants suggested increasing the length of the session to allow for more digital playtime. One participant said:

> *At times there was too much tech talk which meant less time spent using the actual equipment, I think this could be streamlined to make the best use of time in the session.* [Participant 8]

Next, some responses suggested parts of the session were too complex and not pitched at the appropriate level. Participants highlighted that that there was "over-explanation of the technology" (Participant 1), "too much tech talk to start" (Participant 5), and that some parts of the session were "quite confusing" (Participant 4).

Finally, the quality of the internet connection was mentioned by 4 participants as an improvement suggestion, reinforcing that this was the main technology difficulty experienced during the sessions.

Session Relevance and Utility

Following the quantitative questions regarding session relevance and utility, participants were subsequently asked to explain their reasoning in a free-text question. Of the 21 participants, 20 (95%) thought the session was useful and relevant to their future clinical career—the single outlier was "unsure." When asked to expand on their answers, participants gave responses in 4 themes: Digital Future of the NHS, Potential for Health Care Innovation, Impact on Patients, and Ensuring Ideas are Practical.

When considering the relevance/utility of the session, 11 of 21 participants (52%) commented on the Digital Future of the NHS and the need for the workforce to be knowledgeable and prepared:

> *It will become more and more relevant over time.* [Participant 8]
>
> *Realise that tech is coming to the NHS and we need to be prepared to use it in our practice.* [Participant 10]
>
> *Tech is only going to become bigger in the next decade and clinicians need to catch up.* [Participant 3]

Next, 5 participants gave answers that fall under the theme of Potential for Health Care Innovation, recognizing areas for digital integration such as development of virtual patient assessment systems and the interpretation of radiological imaging. The technology still being "in its early stages" (Participant 9), however, was also recognized.

Three participants wrote directly about the impact of technology on patients, which was considered from different angles:

> *Still unsure whether this will benefit patients.* [Participant 9]
>
> *I can see how this type of thing can be used to benefits patients' care in the future.* [Participant 21]

Finally, 3 participants raised the point that that future innovations must be practical. Funding concerns were again

mentioned as well as comments relating to the need to "work out what is realistic" (Participant 12) and the realization that some useful ideas "struggle in their execution" (Participant 2).

Postsession Development

To conclude, participants were asked for their suggestions on how the sessions should be followed up. From the 9 answers provided, 3 themes emerged: Clear Signposting, Focused Technology Support, and Exposure to Technology in Context.

The need for clearly signposted postsession support was raised by 3 participants, to allow ideas and interest generated in the session to be appropriately followed through. One participant talked about the benefit of having a "clear roadmap of steps from this workshop to generating ideas right through to fruition" (Participant 2).

Further, a need for focused technology support was identified by 4 participants, in order to provide more support to participants who had less experience with the technology itself or those who found adapting to the headsets more difficult. An example quote is included below:

> *Would need more time and support if taking this forward as a project.* [Participant 11]

Finally, 3 participants identified that they might benefit from the opportunity to have more exposure to the technology in context, perhaps with opportunities to trial it in clinical simulation or with real patients in the clinical environment.

Discussion

An XR Deep Dive training program has been created for local health care professionals, which has been evaluated as being clinically relevant, successfully increasing local awareness of current digital innovation projects within health care. It is also potentially useful to future clinical practice. This is the first step in developing and enhancing digital literacy and innovation within our health care staff across our integrated care organization.

Session Strengths

Participants indicated that their presession experience of using XR technology was minimal to nonexistent. The integration of digital playtime and first-hand exposure to the technology were reported as being an overwhelming strength of the session. Participants were encouraged to reflect on these practical experiences and engage in collaborative group discussion about potential applications and developments in their own health care settings. This experiential learning is a key component of adult learning theory, where learning takes place in a context-specific cycle of experience, reflection, conceptualization, and experimentation [16]. To provide this all-important context, relevant local use cases of successful XR interventions formed the basis of the practical demonstrations, fueling participants' imaginations of what is achievable within our own organization, thereby lifting the concept of XR integration from an abstract idea to a realistic possibility. For example, the following use cases (developed in-house) were explored (Figure 4):

1. Working with local clinical pain specialists, the Digital Futures team has been able to create a fully immersive tai chi on the beach VR experience (Figure 4A).
2. The successful integration of HoloLens technology to deliver immersive virtual clinics in the patient's homes.
3. How XR technology has been used at TSDFT to develop and deliver interactive empathy (Figure 4B) and patient management training (Figure 4C).

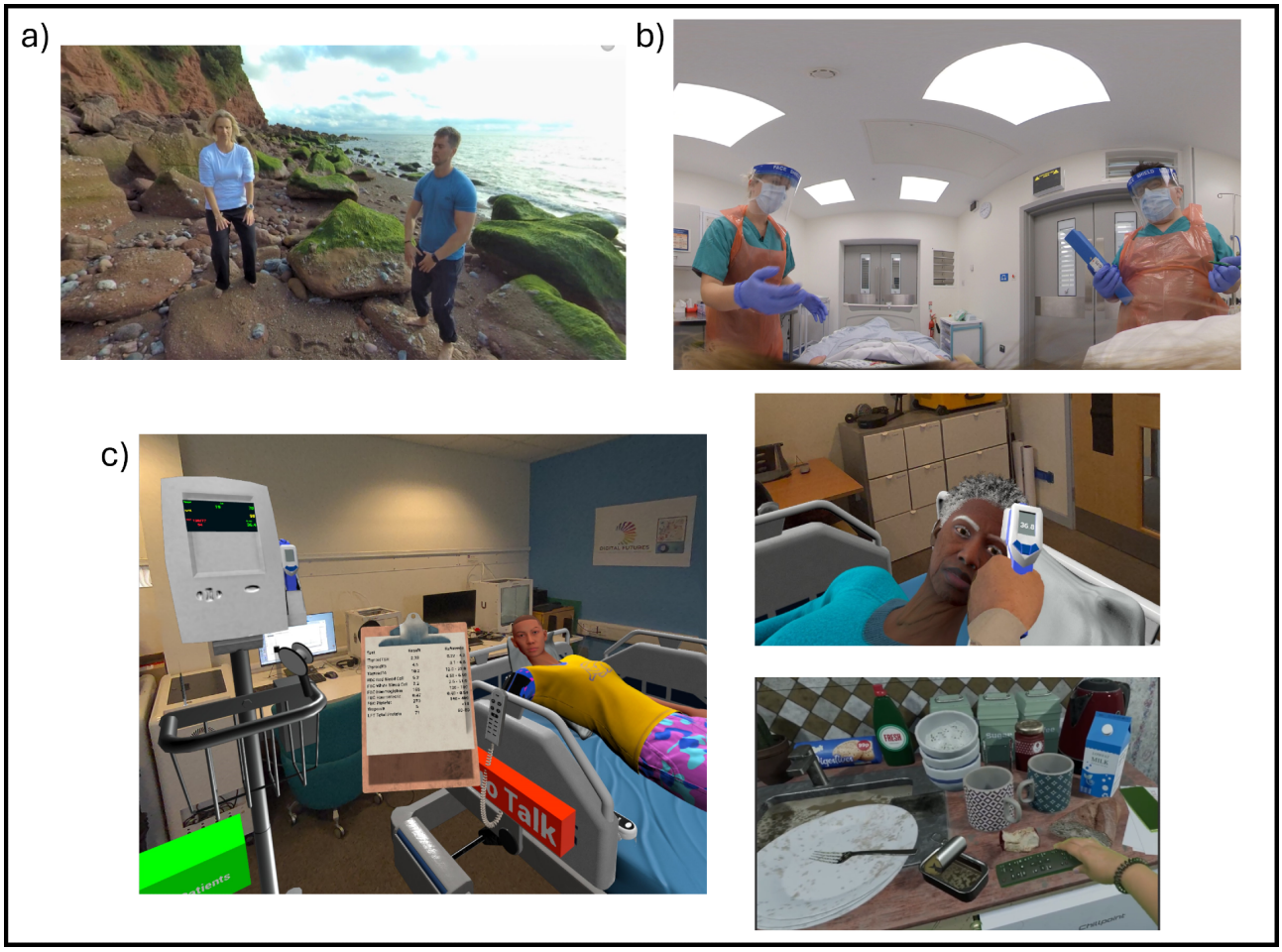
This contextual relevance was another key strength in our survey results, supporting the mantra that "seeing is believing" where emerging technologies are concerned [25].

Significant cultural challenges exist to the widespread adoption of XR technologies across all industries, including feelings of apathy, distrust, confusion, and skepticism [25]. Such cultural barriers are reinforced through our survey, with more than half of respondents exhibiting negative preconceived ideas about the use of XR technology in health care across 3 themes: Patient Concerns, Technology Skepticism, and Cost and Scalability. We believe that such concerns must be addressed head-on by providing staff with the opportunity to experience the technology in action, with time and support to understand its qualities and limitations as well as openly discussing and addressing concerns [25]. After taking part in a Deep Dive session, many participants acknowledged the potential of XR technology for health care innovation and had developed an appreciation of what might be realistically achievable at a local level.

Our co-creation approach to developing digital solutions that are useful and usable in practice was fundamental to the design of the Digital Futures Deep Dive sessions and to addressing these concerns. Having access to a digital expert during the session enabled practical discussions focused on achievable digital goals. Emphasizing cross-fertilization of clinical and digital expertise allows participants to understand that our local Digital Futures Programme aims to produce co-developed, intelligently implemented, and practically driven bespoke patient-focused health care solutions [26], and that digital care transformations are taking place in a positive sphere of negotiation and meaningful dialogue with key stakeholders, rather than being forced upon them [27].

The success of our XR Deep Dive training sessions is encapsulated and demonstrated by a significant number of participants showing active postsession engagement and interest in becoming involved in the local Digital Futures Programme, bringing with them the seedlings of ideas that were sown in the initial XR Deep Dive session.

Figure 4. Examples of local XR use cases demonstrated during the XR deep dive training sessions: (A) immersive VR tai chi on the beach; (B) interactive VR empathy training; and (C) interactive HoloLens patient management training. VR: virtual reality; XR: extended reality.



Areas for Improvement

Our survey revealed that participants felt the balance and complexity of the session could be improved, with respondents requesting more hands-on time with the XR headsets and less presentation time, reinforcing that the strength of the session lies in its integration of practical digital experience. As a result of this feedback, we were able to perform a review of the session design after the first few deliveries and made some intermediate interventions, including increasing the session length from 90 to 120 minutes with more dedicated practical time, streamlining the session presentation, and simplifying the digital-focused background information. This resulted in improved feedback, with improvement comments under the themes of "Improvement of Session Balance" and "Improvement of Session Complexity" occurring far less frequently in the later pilot sessions.

Approximately one-quarter of respondents talked about connectivity issues when asked whether they experienced technology difficulties during the session. Resolution of connectivity issues subsequently became a theme for improvement. Such connectivity issues are unfortunately widespread in the NHS—a survey found that 58% of NHS staff had experienced Wi-Fi blind spots in their trust buildings, and two-thirds agreed that digital innovations in their team had been abandoned due to poor connectivity [26]. This is a limitation of NHS infrastructure and is not within the abilities of this paper's authors to change. However, we recognize—like 98% of NHS staff—that Wi-Fi infrastructure and mobile connectivity are crucial to the future delivery of innovative health care [26] and will therefore continue to play our part in campaigning for improved connectivity as part of our local Digital Futures initiative.

Principal Findings

The Digital Futures Lab is on-site in our NHS trust, and it is bespoke and evolving. It was built to develop and support the digital literacy of all health care staff in our trust. Our evaluation found that most participants came to our training session with no or little knowledge about the use of XR technologies in a health care context or local XR development projects. As expected, most participants had never heard of our new local Digital Futures Programme and were not aware of the investment and facilities available within our own organization recently. Without awareness of the opportunities available, clinical teams simply cannot drive digital innovation. This aligns with the findings of a 2023 survey that lack of digital knowledge and skills within health care teams was considered by three-quarters of surveyed NHS workers to be a significant barrier to innovation [26]. Furthermore, this emphasizes the Topol recognition that a culture of NHS digital innovation can only be achieved when coupled with a learning culture that supports frontline staff to explore new technologies and the opportunities they present for patient care [1].

An integral component to building such a culture is having a cohort of learners who are motivated to explore the opportunities presented by advancing digital technologies. An appetite to explore and embrace digital advancements to transform patient care has been identified among health care workers on a national scale [26]. Similarly, many participants in our survey acknowledged the upcoming digital age of the NHS, recognized digital technology as a tool for innovation, and cited their curiosity to learn about such innovation opportunities as motivation for seeking digital training.

Overall, in our local health care workforce, there exists a knowledge gap regarding XR potential and current local opportunities coupled with a strong desire to rectify this, indicating a clear need for the XR Deep Dive training session we have created. After taking part in our session, all respondents reported an increased awareness of local digital innovation and most felt inspired to get involved in future digital projects themselves, highlighting that our sessions have been successful in meeting their aims.

Future Directions

Feedback from pilot sessions has supported the need for our new XR Deep Dive training sessions and has informed the refinement of the original session design as part of a quality improvement cycle. Intermediate interventions to address initial concerns regarding session balance and overcomplexity have already been successfully implemented, and there remains scope for further improvement. For example, future directions of the XR Deep Dive training program may involve a tiered approach to cater for participants of different starting abilities and experiences, potentially incorporating "beginner," "intermediate," and "advanced" training sessions, which can be accessed either in isolation or as a progressive series. Future evaluation of such an expansion of the training program would offer further insights into how we can successfully fulfill the NHS Long Term Workforce Plan of upskilling and training staff in our NHS trust to maximize digital technologies to improve health care delivery for the benefit of patients locally [3]. Future research will also inform us about the different technology behaviors of individuals and help us develop insights on how behavior change can be encouraged.

Digital transformation, and XR health care technologies in particular, are rapidly evolving and driving change. Maturation of hardware and software means content is becoming more sophisticated, user friendly, and seamlessly integrated into the real world [25]. Training programs—such as the one we have developed—will therefore also be required to evolve. Regular periodic reviews of the session content must be scheduled with updates as required, to ensure the training does not become outdated and irrelevant. Further, as use of technology in our local trust increases, the use cases demonstrated in the XR Deep Dive training sessions must also be reviewed to ensure they remain current and engaging. Showcasing use cases tailored to the participants' own context will become easier as more local health care specialties adopt XR innovation.

As the training program grows, we must ensure its sustainability. This will involve the recruitment of local "clinical digital champions"—as identified in Topol [1]—to deliver peer-to-peer training, sharing their knowledge and unique experiences. Recruitment and training of digital experts must also be maintained—and increased proportionately—at the trust level. Ongoing funding must be secured in line with the program growth, which will require a funding strategy as part of the wider Digital Futures Programme in our trust. A robust and sustainable follow-up support model must be established to bridge the gap between this initial training session and adopting XR solutions in the clinical environment. Sparking the imagination of what is possible in the realm of local XR health care technology is trivial if participants do not subsequently have access to the technical support and expertise required to conduct trials within their own clinical spaces. We have already begun to tentatively explore a model of "Digital Clinics" for this purpose, but data from our survey emphasize how follow-up support must be focused, context-specific, and clearly signposted. Refining a sustainable follow-up model that meets these criteria is the next step in the development of this training program.

Finally, digital health care transformation is certainly not without its ethical challenges, including concerns around access, consent, inclusivity, privacy, and dignity [1,28]. As digital innovation training evolves, it must incorporate these ethical discussions and continue to tackle cultural barriers. Encouraging honest and open dialogue will be key to finding workable local solutions to ethical challenges and ensuring a true co-design culture is adopted. Our survey highlights staff concerns that XR technology will remove the personal aspect from patient-clinician relationships, thereby dehumanizing care. This concern is also recognized in the Topol review. Our local Digital Futures goal aligns with that of Topol: to focus on how digital technologies can enhance, rather than retract from, our human interactions. We are proud that our local digital projects prioritize the humanistic aspects of care and have built our training to showcase this. As digital innovation and the associated awareness training evolves, we must not lose sight of our core values.

Limitations of This Paper

This paper explores a small, single-center pilot of a new local training intervention. Its findings are intended to inform future directions in our own trust and may not be generalizable to a wider context.

First, given the voluntary, self-selection sampling used to recruit participants to the Deep Dive pilot sessions, it is likely that our survey suffers from selection bias, capturing the views of staff who were already motivated to undertake the training in the first place. Given that a significant number of survey respondents talked about a prior interest in technology and a curiosity to explore new digital opportunities further as a reason to sign up to the pilot sessions, it is likely that our data do not capture the cohort of staff in our trust who are true digital skeptics. To obtain a wider spectrum of opinions, for future iterations of this pilot, we should aim to recruit staff members who do not have prior motivation for engaging in digital training sessions. This will provide insights into how we can effectively engage digital-skeptic staff to engage in the technology

advancements being implemented both in our local trust and nationally within the health service.

Second, feedback was collected via an online feedback form accessed via a QR code at the end of the session. Not all session participants completed the feedback (60% response rate), possibly owing to the fact there was no physical form and they never got around to submitting it online. Concerns around nonresponse bias must therefore be considered when interpreting our findings. Obtaining feedback online is an established challenge [29]. To ensure a more complete representation of participant views in future, it may be preferable to supplement a feedback form with a recorded feedback focus group at the end of future sessions.

Conclusion

Having identified a gap in real-world working models of health care workforce XR awareness and development training, we have designed and implemented XR Deep Dive training sessions for health care staff. This was one of the principle aims of our Digital Futures Programme. These sessions provide contextually relevant XR technology awareness training and are the first step in working toward the goal of nurturing digitally literate health care workforces who have the knowledge and skills to embrace transformative technology in the improvement of patient care, as per Topol [1]. Our session design draws on Experiential, Active, and Contextual Learning theories by showcasing local use cases of the technology in practice, prioritizing hands-on digital playtime and emphasizing the vital cross-fertilization of clinical and digital expertise in the co-creation of digital solutions that are useful and usable in practice. Data from the pilot sessions suggest that we have created a training session that is engaging as well as relevant and useful to future clinical practice. The results from this paper will help to inform future directions for developing digital awareness training in our trust.

Acknowledgments

We thank Dr Matthew Halkes and Mr Christopher Matthews, who have contributed to the development of the Digital Futures Programme at TSDFT. Funding for the development of the Digital Futures Lab at TSDFT was provided by Health Education England and TSDFT League of Friends. This work would not have been possible without support from TSDFT. Artificial intelligence was not used when writing this manuscript.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CG wrote the manuscript and carried out thematic analysis of results. PG carried out thematic analysis of the results. JRL, NP, CG, and JW were involved in the design and delivery of the Deep Dives training program. JRL and NP secured funding and originally developed the "Digital Futures: Human Centred Digital Innovation" program and set up the Digital Futures Lab at TSDFT. All authors reviewed and edited the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire survey used to collect participant feedback following pilot XR Deep Dive training sessions.
[PDF File, 483 KB - xr_v2i1e57361_app1.pdf]

References

1. Topol E. The Topol Review: Preparing the healthcare workforce to deliver the digital future. An independent report on behalf of the Secretary of State for Health and Social Care. : NHS; 2019 URL: https://topol.hee.nhs.uk/the-topol-review/ [accessed 2025-02-16]
2. Morris D, Solon L, Simpson J, Read O, Abel J. Tech powered healthcare: a strategic approach to implementing teachnology in health and care. PwC. 2020. URL: https://www.pwc.co.uk/healthcare/assets/documents/making-healthcare-digital-tech-powered-health.pdf [accessed 2025-02-16]
3. The NHS long term workforce plan. 2023. URL: https://www.england.nhs.uk/wp-content/uploads/2023/06/nhs-long-term-workforce-plan-v1.21.pdf [accessed 2024-02-10]
4. Digital technologies in healthcare: a systematic review and bibliometric analysis. Int J Onl Eng 2022;18(8):34-48. [doi: 10.3991/ijoe.v18i08.31961]
5. Kilkelly F, O'Brien R, Ticho S. The growing value of XR in healthcare in the United Kingdom. 2021. URL: https://www.xrhealthuk.org/the-growing-value-of-xr-in-healthcare#Appendix [accessed 2025-02-16]
6. Yeung AWK, Tosevska A, Klager E, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. J Med Internet Res 2021 Feb 10;23(2):e25499. [doi: 10.2196/25499] [Medline: 33565986]

7. Southworth MK, Silva JR, Silva JNA. Use of extended realities in cardiology. Trends Cardiovasc Med 2020 Apr;30(3):143-148. [doi: 10.1016/j.tcm.2019.04.005] [Medline: 31076168]
8. Zhang J, Lu V, Khanduja V. The impact of extended reality on surgery: a scoping review. Int Orthop 2023 Mar;47(3):611-621. [doi: 10.1007/s00264-022-05663-z] [Medline: 36645474]
9. Dadario NB, Quinoa T, Khatri D, Boockvar J, Langer D, D'Amico RS. Examining the benefits of extended reality in neurosurgery: a systematic review. J Clin Neurosci 2021 Dec;94:41-53. [doi: 10.1016/j.jocn.2021.09.037] [Medline: 34863461]
10. Pereira MF, Prahm C, Kolbenschlag J, Oliveira E, Rodrigues NF. Application of AR and VR in hand rehabilitation: a systematic review. J Biomed Inform 2020 Nov;111:103584. [doi: 10.1016/j.jbi.2020.103584] [Medline: 33011296]
11. Wenk N, Buetler KA, Penalver-Andres J, Müri RM, Marchal-Crespo L. Naturalistic visualization of reaching movements using head-mounted displays improves movement quality compared to conventional computer screens and proves high usability. J Neuroeng Rehabil 2022 Dec 9;19(1):137. [doi: 10.1186/s12984-022-01101-8] [Medline: 36494668]
12. Duarte ML, Santos LR, Guimarães Júnior JB, Peccin MS. Learning anatomy by virtual reality and augmented reality. a scope review. Morphologie 2020 Dec;104(347):254-266. [doi: 10.1016/j.morpho.2020.08.004] [Medline: 32972816]
13. Dyer E, Swartzlander BJ, Gugliucci MR. Using virtual reality in medical education to teach empathy. J Med Libr Assoc 2018 Oct;106(4):498-500. [doi: 10.5195/jmla.2018.518] [Medline: 30271295]
14. Digital Futures: Human Centred Innovation. Torbay and South Devon NHS Foundation Trust. 2024. URL: https://immersive.tsdft.uk/ [accessed 2024-02-10]
15. Morrison M. What is the IDEO Deep-Dive. Rapid BI. 2018. URL: https://rapidbi.com/deep-dive-brainstorming-technique-ideo/ [accessed 2024-07-21]
16. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development: Prentice Hall; 1984.
17. Mitchell A, Petter S, Harris AL. Learning by doing: twenty successful active learning exercises for information systems courses. JITE:IIP 2017;16:021-046. [doi: 10.28945/3643]
18. Coulter A, Oldham J. Person-centred care: what is it and how do we get there? Future Hosp J 2016 Jun;3(2):114-116. [doi: 10.7861/futurehosp.3-2-114] [Medline: 31098200]
19. Johnson EB. Contextual Teaching and Learning: What It Is and Why It's Here to Stay: Corwin Press Inc; 2002.
20. Pollock PH, Hamann K, Wilson BM. Learning through discussions: comparing the benefits of small-group and large-class settings. Journal of Political Science Education 2011 Feb 8;7(1):48-64. [doi: 10.1080/15512169.2011.539913]
21. Martin HJ. Improving training impact through effective follow‐up: techniques and their application. Journal of Management Development 2010 Jun 15;29(6):520-534. [doi: 10.1108/02621711011046495]
22. Self selection sampling. Laerd Dissertion. 2012. URL: https://dissertation.laerd.com/self-selection-sampling.php#first [accessed 2024-07-21]
23. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf 2016 Dec;25(12):986-992. [doi: 10.1136/bmjqs-2015-004411] [Medline: 26369893]
24. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology: American Psychological Association; 2023, Vol. 2:57-71.
25. Seeing is believing: how virtual reality and augmented reality are transforming business and the economy. PwC. 2019. URL: https://www.pwc.com/gx/en/technology/publications/assets/how-virtual-reality-and-augmented-reality.pdf [accessed 2024-02-10]
26. Mind the gap! Delivering digitally enabled change to the NHS. BT healthcare. 2023. URL: https://pages.business.bt.com/rs/407-ICJ-312/images/aspectus_bt_health_whitepaper.pdf [accessed 2024-02-10]
27. Houghton E, Baczor L. Workplace technology: the employee experience. CIPD. 2020. URL: https://www.cipd.org/globalassets/media/knowledge/knowledge-hub/reports/workplace-technology-2_tcm18-80853.pdf [accessed 2025-02-16]
28. Evans J. The IEEE Global Initiative on Ethics of Extended Reality (XR) report: Extended Reality (XR) Ethics in Medicine. : IEEE; 2022 URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9714059 [accessed 2025-02-16]
29. Nulty DD. The adequacy of response rates to online and paper surveys: what can be done? Assessment & Evaluation in Higher Education 2008 Jun;33(3):301-314. [doi: 10.1080/02602930701293231]

Abbreviations

AR: augmented reality
NHS: National Health Service
TSDFT: Torbay and South Devon Foundation Trust
VR: virtual reality
XR: extended reality

Edited by T Leung; submitted 14.02.24; peer-reviewed by TW Yu, T Davidson; revised version received 12.10.24; accepted 14.10.24; published 27.02.25.

Please cite as:
Galvin C, Watt J, Ghatnekar P, Peres N, Rees-Lee J
A Local Training Program to Increase Awareness of Emerging Extended Reality Technologies Among Health Care Professionals: Development Study
JMIR XR Spatial Comput 2025;2:e57361
URL: https://xr.jmir.org/2025/1/e57361
doi:10.2196/57361

Role of Augmented Reality in Tertiary Care: Qualitative Investigation Using Thematic Analysis

Jacob Hobbs[1], BSc; Christopher Bull[1], PhD; Caroline Claisse[1], PhD; Mat Elameer[2], MBBS; Richard Davison[3], PhD

[1]Open Lab, School of Computing, Newcastle University, Floor 1, Urban Sciences Building, Newcastle Helix, Newcastle upon Tyne, United Kingdom

[2]Stroke Research Group, Institute for Ageing, Faculty of Medical Sciences Newcastle University, Newcastle upon Tyne, United Kingdom

[3]Games Lab, School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom

Corresponding Author:
Jacob Hobbs, BSc
Open Lab, School of Computing, Newcastle University, Floor 1, Urban Sciences Building, Newcastle Helix, Newcastle upon Tyne, United Kingdom

Abstract

Background: While augmented reality (AR) as a concept is not new, it is still an emerging technology with a wide range of applications that it could provide value for. In the medical field, AR is becoming ever more prevalent, but while it has been applied to various medical tasks, it is far from commonplace. Radiological imaging has been suggested as one of these applications, and the radiology workflow capacity crisis the United Kingdom's National Health Service is experiencing is a potential opportunity for technology to alleviate pressure. Understanding clinical stakeholders and current systems is important for identifying design opportunities for developing AR to enhance interactions and gain more from radiological images.

Objective: This study had 3 key aims. First, to build an understanding of the field in the context of AR; second, to understand the stakeholders and workflows surrounding radiological images; and finally, to suggest how AR could integrate within these workflows and current practices in order to provide value.

Methods: We conducted 14 interviews with hospital-based consultants in a range of specialties and then completed a thematic analysis on the transcripts in order to find trends that suggest what value AR could add to radiological imaging, where that value could be added, and who would benefit. We implemented reflexive thematic analysis to develop themes from across the interviews, which were then built on to suggest design implications.

Results: We find that the need for efficiency in image evaluation is present across many roles, regardless of the clinical question, but consultants can be resistant to new technology. Additionally, we find that the current capability of AR technology could be of greater benefit to radiologists as opposed to surgeons or other practitioners. We discuss these findings for the development of AR applications and present 3 design implications that stand as our core contribution.

Conclusions: We conclude with 3 design implications for the application of AR within radiological imaging based on the results of our thematic analysis and frame them within the Human-Computer Interaction and medical fields. The first design implication highlights efficiency and how AR has the potential to allow for quicker comprehension and measurements. Second, we suggest that the capability of AR tools should complement existing techniques and not simply replicate current ability in 3 dimensions. Finally, the integration of AR tools with existing workflows is crucial in the uptake of the technology in order not to negatively disrupt practice.

(JMIR XR Spatial Comput 2025;2:e68810) doi:10.2196/68810

KEYWORDS

augmented reality; clinical practice; radiology; surgery; thematic analysis

Introduction

Overview

Augmented reality (AR) for clinical use was first mentioned in 1982 [1] and 1992 with a head-mounted display (HMD) [2]. Using AR in a clinical setting is not a new concept, but it is still in relative infancy [3,4] with many suggestions as to the specific applications [4-6]. It is a promising application area of AR with many examples presented [7]. Despite this, it is still an emerging technology, and there is very little uptake of AR in the day-to-day of clinical practice [6]. The motivation of this study is to explore where this emerging technology could provide value to modern medical practices, specifically radiological imaging. The Royal College of Radiologists highlights the urgent workflow capacity crisis in terms of the number of staff not keeping pace with the increasing demand for imaging. Increased strain is therefore placed on existing staff within the National Health Service (NHS), the publicly funded health care system in the United Kingdom [8]. By exploring these problems

in the context of AR, we can begin to understand how the technology could fit into the goals or requirements that are present in today's practice, such as increased efficiency or higher accuracy.

In this study, we conducted an interview study to investigate the current clinical landscape of radiological imaging in modern medicine to better understand the potential roles AR could play and the value it could bring. This was achieved by conducting a set of interviews with consultants in different tertiary care specialisms (highly specialized care) aimed at exploring the current practices and perspectives of professionals who work with radiological images, in the context of using AR technology. As a result, we propose 3 design implications to consider when designing AR systems for clinical use, which stand as our core contribution. Our design implications were informed by experiential accounts and opinions regarding what radiological images are used for, how they are used, and what stakeholders gain from them. We recruited surgeons and radiologists as key stakeholders, and a focus was put on the interactions these stakeholders had with the images used during clinical workflows. This enabled us to examine the contents of the interactions as well as the users' experiences and opinions on how successful they were in the context of looking for opportunities to design for AR. The current tools used, how the tools are integrated into practice, and opinions on them were also considered.

The aim of these interviews was threefold: to gain an understanding of the field in the context of this technology, to gain an understanding of the stakeholders and workflows surrounding medical images, and to begin to understand the role that AR could play within these workflows. The interviews were semistructured around questions that sought to clarify medical facts, explore the opinions and discrepancies of current practice, while also probing attitudes toward the problems, opportunities, and new technologies that are faced. The interviews have been analyzed using reflexive thematic analysis [9,10] to understand trends and contradictions across the data set. This analysis is intended to understand what value AR could provide in a clinical environment and, therefore, identify application and interaction design opportunities and suggest some design implications. Going forward, this will allow us to begin to identify some of the needs of tertiary care practitioners in the context of this technology. The contribution of this work is the empirical understanding gained through the thematic analysis and the 3 design implications developed based on this analysis. The thematic analysis aims to understand the needs and challenges experienced by hospital-based consultants, and the design implications are developed through and justified by this thematic analysis.

Background

Development of AR

AR superimposes digital objects into the users' view in real-time using a headset or other device. The aim is to add virtual components to the user's field of view to provide them with additional information while carrying out a task [5]. Although the term was coined in 1992 [11], the technology has seen a boom in interest in recent years [12]. It was at this early stage in 1992 that AR would be suggested as a tool to aid surgery by Rosenburg [2]. Rosenburg suggested that just as a physical ruler can be used as a tool to aid in drawing a straight line on a piece of paper, AR could be used to guide surgeons' incisions, and that AR would be better than any physical tool for this task as the virtual components could be partially submerged in the anatomy to strictly follow key lines and boundaries.

Since this time, AR technology has developed with advancements such as viable HMDs, allowing wider and more creative adoption [13]. There is little clinical use of AR, but interest in the technology for use in this space is growing [6]. It has been suggested for image-guided surgery (IGS), as Rosenburg did, but also for tasks such as medical training, clinical psychology, diagnostics, surgical planning, and rehabilitation [5,14,15].

HMDs are the dominant way of using AR, and technological developments have meant that they can display content accurately enough to enable convincing interactions. However, technological and usability issues persist around AR HMDs [16-18], with the effectiveness and accuracy of AR in many clinical tasks difficult to validate and therefore remaining to be proven [19,20]. A key set of issues documented across a variety of AR headsets is the perceptual inaccuracies and issues that can arise. Perceptual issues are an important area of research, as regardless of the domain or individual application, an otherwise perfect AR experience could be made intolerable by physical symptoms as a result of inaccurate perceptual cues. This is particularly true in a medical environment where the accuracy of the tools used can have an implication on a patient's life [21].

Poor perceptual cues can place stress on a user, resulting in symptoms such as motion sickness, nausea, and visual fatigue. Focal rivalry is a common example of inaccurately represented virtual content, placing unmanageable stress on the users' vision. Focal rivalry is where the eyes cannot focus on 2 objects at different depths at the same time and therefore have to switch between focusing on the physical object and the virtual, a requirement rarely seen in the natural environment [22] .

The vengeance-accommodation conflict is another common perceptual issue that has been documented to cause physical symptoms. The vengeance-accommodation conflict [23] is caused by the eye's 2 mechanisms of focusing competing against one another. Most modern HMDs have a fixed focal depth of around 2 m, but as virtual content is moved away from this plane, inaccurate depth cues are created, often out of the bounds of what a user's eyes can tolerate [24].

Gold Standard: AR IGS

AR IGS was one of the first clinical applications AR was suggested for, and is still a key area of interest in medical AR research, and is a clear application of the technology [19]. It can be argued that AR IGS is the gold standard of clinical AR as there is broad agreement that having live guidance for operations would be of significant value to the surgeon, resulting in a higher chance of successful surgery [15,25,26]. The theoretical implementation of IGS is that guides such as 3D virtual representations of anatomy, built from preoperative

scans, are overlaid onto the patient in order to allow the surgeon to see anatomy below the surface and more easily identify structures, as well as the boundaries between them. This is intended to speed up procedures, reduce trauma, and reduce recovery time [27].

However, significant issues remain with reaching this goal, which can be broadly divided into technical and usability issues. An important technical issue is registration, the process of aligning virtual components with their physical counterpart. Registration requires enough identifiable points, which can be known as markers, to be present on both the virtual object and the physical anatomy in order to map one to the other, and in a lot of cases, there are not enough. Machine learning algorithms have been used to approach this problem and generate nets of points across both objects, then map them together [28]. Bertolo et al [29] cite registration as a prominent unsolved challenge and state that in the era of "precision surgery," clinicians will expect error margins to be negligible.

In addition to the technical issues, it is still unclear how best to present virtual content to a surgeon for IGS. Dilley et al [30] suggest that even with perfect registration, surgical performance is reduced when virtual content is overlaid onto the surgical site. Their work suggests that even in a currently fictional environment where perfect registration can be achieved, projecting the images used for guidance beside the patient, unregistered, provides a better outcome.

Determining the best way to present virtual content is one of many usability issues that remain unsolved. Successfully determining what virtual content is best to display to a surgeon can only be useful if the methods the surgeon uses to interact with the content are intuitive, unobtrusive, and effective. The study by Eddie [19] suggests that the visualization and interaction challenges are the biggest challenges facing AR surgical guidance.

AR IGS is likely to provide significant value to surgeons once its value and accuracy can be proven. However, there are multiple issues that all need to be overcome to achieve this. IGS is far from the only application of clinical AR to provide value [31].

Modern Clinical Applications

Modern clinical applications of AR can broadly be split into 3 categories: intraoperative (eg, AR IGS discussed above), education and training, and presurgery tasks. The educational and training applications of AR are very broad, ranging from using AR to facilitate the learning of anatomy to safer, more repeatable surgical training [32]. AR has the potential to provide more immersive, repeatable, readily available training and education in the medical field, allowing everyone from medical students to qualified surgeons to take in new knowledge in a new way [4]. In situations where a qualified surgeon is learning a new procedure, AR allows a safer, no-pressure environment for the surgeon to understand how the procedure works and repeatedly practice the intricacies [33].

There are several applications of AR in the presurgery domain, principally, diagnostics and surgical planning. AR for surgical planning allows the surgeon to view preoperative images such as computed tomography (CT) and magnetic resonance imaging (MRI) images, as 3D models of the surgical site before the procedure [15,34]. This is suggested to allow the surgeon to gain a better understanding of the surgical site and relationships between structures, meaning they can plan how a procedure will be approached and be more prepared for potential complications [35].

Douglas et al [36] suggest that using AR could improve diagnostic accuracy and speed up the diagnostic process when viewing cross-sectional images such as CT and MRI. Pelargos et al [37] state that "surgical planning is inherently a 3D task" and that virtual reality and AR technologies could help by improving the understanding of the complex anatomical relationships. These tools have the potential to offer better visualization of areas of interest and therefore improve the understanding and the speed at which decisions can be made [34,38]. Trestioreanu et al [39] argue that AR and virtual reality have the potential to improve radiology health care by improving the cognitive experience, by reducing the cognitive load that a clinician undergoes when viewing 2D slices of 3D anatomy. They go on to suggest that while a few 3D visualization methods currently exist, they do not offer the increased practicality or ergonomics that AR approaches could offer.

As it stands, there is very little AR in day-to-day clinical practice [6,19]. The literature discussed above has directed our work to focus on investigating where AR could be applied in the presurgical domain around radiological images and what value the technology could bring. This is a promising area of research where AR technology could be harnessed effectively. Our work is positioned to direct future research and contributes to the body of literature directing the development of AR applications for radiology, based on expert end user experiences.

Methods

Ethical Considerations

This work was granted ethical approval by Newcastle University ethics committee (27432/2022). Participants gave their informed consent to the interviews, and it was made clear that they could withdraw their participation at any time. Ages of participants were captured as ranges and demographic information captured was kept to a minimum to maintain participant privacy. Participants received no compensation for their time.

Recruitment Process and Participants

For this study, 14 semistructured interviews were conducted with medical professionals from a range of specialties to enable us to determine how practices and perspectives around radiological images vary across specialisms and hospitals. Five of these interviews were with radiologists with various subspecialties, while the remaining 9 were with other consultants in areas such as cardiology, cardiothoracic surgery, general surgery, orthopedic surgery, and clinical oncology. Participant demographic details are summarized in Table 1. All of the participants were male, which is acknowledged and discussed in the Limitations section. Demographic questions were voluntary, and as such, some participants chose not to share some personal information, which is denoted in Table 1 with

"—." The participants worked at 8 different hospitals, 5 of which were in the Northeast and Northwest of England. Two of the remaining were London hospitals, and one on the South coast of England. Initial participants were recruited through the authors' host university medical school via public staff lists. These participants were then asked to refer other potential participants, especially from other hospitals and regions of the United Kingdom, having a snowballing effect. One of the participants was previously known to the researchers, 2 participants were recruited through mutual acquaintances, and all others were previously unknown to the researchers.

Table . Participant demographic information.

ID	Age range (years)	Ethnicity	Role	Time in current role
A	45 - 54	White British	Consultant interventional cardiologist	19 years
B	45 - 54	Mixed White Asian	Consultant cardiologist	11 years
C	45 - 54	Indian	Consultant cardiologist	14 years
D	__[a]	—	Consultant oncologist	—
E	55 - 64	White British	Cardiac surgeon	20 years
F	55 - 64	—	Consultant interventional and diagnostic neuroradiologist	—
G	45 - 54	White British	Thoracic surgeon	11 years
H	—	—	Orthopedic surgeon	—
I	45 - 54	White	Consultant general surgeon	10 years
J	55 - 64	White British	Cardiothoracic surgeon	5 years
K	25 - 34	Mixed White Arab	Consultant neuroradiologist	9 months
L	35 - 44	White British	Consultant radiologist (nuclear medicine)	10 years
M	35 - 44	Indian	Consultant radiologist	4 years
N	35 - 44	White British	Consultant cardiothoracic radiologist	4 years

[a]Not available.

Interview Process

Semistructured interviews were chosen over fully structured interviews in order to be more open-ended and allow greater flexibility for free conversation. The interviews were all conducted over Microsoft Teams (Microsoft Corp) and lasted between 30 minutes and an hour. Fourteen questions were drawn up based on prior reading in the area, in context with the aims of the interviews. The first objective of the interviews was to act as a means of gaining knowledge of relevant medical specialisms, their current working practices, and collaboration methods across NHS trusts. This way, the authors could build a solid base of knowledge of the field that allowed an appreciation of the context and the identification of nuance in practice. The current practice surrounding radiological images was a key point here. This included establishing how images are used, the tools used to interact with them, how the tools and requirements change between different specialties, and what is gained from the images themselves, that is, what questions they are used to answer. This continued into establishing the current workflows around these images, the communication between stakeholders in reference to imaging, particularly the communication between these hospital-based consultants, how information flows between stakeholders, and what this process looks like from a patient's perspective.

The clarification of this base knowledge laid the groundwork for more in-depth questions exploring the opinions around these areas: how useful the tools are, how the tools vary, and how personal preference influences both the use of tools and the practice itself. This was then followed by questions about their experience level, their use, and the utility of AR, which were intended to explore the current uptake of this technology and opinions on AR as it exists at the current point. Finally, there were questions about the future of the participant's specialty and what technologies they saw as having a notable impact.

Analysis Process

Overview

The interviews were recorded and transcribed, providing 14 transcripts that could then be subjected to reflexive thematic analysis. This allowed the authors to establish trends and reveal insights across the whole interview dataset. Thematic analysis is a set of methods for data analysis to develop, analyze, and interpret patterns across a qualitative dataset. Reflexive thematic

analysis, developed by Braun and Clarke [40], is an interpretive qualitative approach that encourages critical reflection of the role the researcher plays in the analytic process and their research practice. Braun and Clarke talk about the inherent presence and necessity of biases and how they are integral to reflexive thematic analysis. Reflexivity is integral to this analysis method, "We must question why we think what we think. Bias, prior knowledge and who we are shapes subjectivity" [41]. Thematic analysis is a set of interpretive qualitative analysis methods, and as such, the researchers' perspectives and biases are used as tools for analysis. It is important to understand these biases when carrying out this form of analysis in order to understand the context used to come to conclusions and how that context influences the conclusions. It is at this point that we, as authors, must consider our positionality.

Positionality Statement

We cannot expect the interviews to uncover the full range of opinions and practices within a particular medical specialization, but we aim to include a sufficient range of participants in order to be representative of the area. Where this is not possible, the researchers acknowledge which groups could not be recruited and the effect this may have on the analysis. Authors 1 and 2 (JH and CB), who conducted the analysis, are computer scientists in Open Lab, a Human-Computer Interaction (HCI) laboratory in the School of Computing at Newcastle University, United Kingdom, with experience in digital health, but no formal medical training. Our expertise lies in qualitative methods and designing technologies for specialist user groups. The remaining authors contributed and provided additional context after the analysis was completed. These authors can be considered tech-savvy and protechnological innovation, which will lend the interpretations of the analysis to following this philosophy. Other researchers will bring different perspectives and have different experiences informing their analysis and will, therefore, come to different conclusions.

Implementing Reflexive Thematic Analysis

Thematic analysis is an overarching term for a flexible set of methods designed to interrogate qualitative data. This study follows the updated version of reflexive thematic analysis by Braun and Clarke [10], which builds on their original work [40]. It is appropriate for this study as the aim of the analysis was to understand the common themes and contradictions across all 14 interviews in order to provide rich insights across a range of specialties [42]. A reflexive approach was applied to this study to foster an organic coding process and to use researcher subjectivity as a tool. This approach means that "themes cannot exist separately from the researcher—they are generated by the researcher through data engagement" [42] and is a direct result of researcher subjectivity being positively exploited. While thematic analysis is a theoretically flexible set of methods, it is important to understand the theoretical base and assumptions being brought to the analysis [9]. For this study, the authors approached the analysis from a relativist ontological position and used a constructionist epistemology. This means that the authors could explore the meaning from the participants in context and be directed by this, constructing meaning and evidence through the analysis. This is opposed to a more traditional realist postpositivist approach, where it is considered that a single objective truth exists within the data, and it is the researcher's job to find it [10,43].

As defined by Braun and Clarke [10], an inductive coding process was used in this study. This was to enable the focus to be put on the participants' experiences and opinions, and as such, allow themes and contradictions between participants to be brought to the surface. As previously mentioned, this inductive process was colored by inherent epistemological and ontological assumptions as "you cannot enter a theoretical vacuum when doing TA" [44]. In a similar vein, a combination of both semantic and latent codes was used throughout the coding process. The semantic codes captured the explicit, surface-level detail that was being communicated while the latent codes grasped the deeper, more implicit points being made. This combination allowed for a thorough and meaningful analysis.

In terms of the analytic process, for reflexive thematic analysis, Braun and Clarke [10] detail 6 phases: familiarization, coding, initial theme generation, developing and reviewing themes, refining, defining and naming themes, and writing up. The familiarization phase was achieved in 2 ways, first, with the lead author conducting the interviews, there was an initial exposure to all of the data in the context it was given. Second, through the transcription process. Automated tools were used for the bulk of the transcription, but the lead author checked each transcript against the interview recording. This ensured that the transcripts were accurate while also contributing to the familiarization phase of the analysis. The coding and theme generation were primarily carried out by the lead author, with the second author offering opinions and challenging decisions after each round. Two full coding rounds were completed, and theme generation was completed over 3 iterations with the second author contributing opinions after the initial coding of 2 transcripts, again after all transcripts had been coded and between iterations of theme development. This contributed to the robustness of the coding and theme generation phases, as it was an opportunity for biases and assumptions to be questioned. The second author contributed to the analysis by reviewing initial codes and themes and probing into the reasoning behind them. This provoked further reflection on the codes and themes throughout the analysis process and meant that assumptions could be challenged, resulting in a deeper meaning being developed. Initial coding resulted in several hundred codes, but upon review, in between and after each coding round, similarities between codes were identified, and clustering codes together allowed for easier interpretation for theming. We then initially grouped codes into 12 broad patterns (eg, multidisciplinary teams [MDTs], personal preference, increased reliance on imaging, and relationships with imaging), which could then be reviewed between authors and the logic or biases challenged. These were then iterated on with the context and theoretical positioning discussed above to develop the themes presented below. Each theme articulates a different aspect of the conversations had while sitting within the context of this work.

Results

Overview

The results are presented as the 4 themes developed through the thematic analysis process. These 4 themes are that communication is largely verbal or written, which acknowledges observations around how communication is conducted regarding radiological images and how it is mostly via the radiologists' report and in MDT meetings. Inconsistencies and personal preference in practice encapsulate the extent to which personal preference and other choices change practice. Extended reality (XR) maturity for surgery covers the opinions of current XR technology, AR in particular, and how there is potential for it in certain areas of practice, but there are still significant issues preventing the mass uptake. Finally, increased reliance on imaging is a known issue in radiology, but this theme explores the opinions in this area and the potential ramifications interviewees believe they will experience. These themes are summarized in Table 2.

Table . Theme table summarizing themes and characteristics.

Theme	Subthemes	Characteristics
Communication is largely verbal or written	• MDTs[a] • Written reports • An intuition of knowing what questions the next clinician will ask	MDTs and radiologists' written reports are stored and interacted with via PACS[b]. MDT communication is high-level, aiming to reach decisions quickly.
Inconsistencies and personal preference in practice	• Discrepancies in reporting • Discrepancies in tools used	Tools, expertise, and practice vary between consultants, departments, and trusts. Given the same tools, slightly different results are likely to be reached.
XR[c] maturity for surgery	—[d]	Current technological state of the art. The potential impact of AR[e] technology.
Increased reliance on imaging	• Efficiency required • Acceptability of new technology	Efficiency is a big concern for everyone, but particularly radiologists. The appeal of new technology to clinicians varies—value must be proven.

[a]MDT: multidisciplinary team.

[b]PACS: picture archiving and communications system.

[c]XR: extended reality.

[d]Not available.

[e]AR: augmented reality.

Communication Is Largely Verbal or Written

Including a variety of hospital-based consultants as participants in this study provided insight into the communication between these 2 parties and how radiological images are used in this process. Two of the important opportunities for communication in terms of radiological images are the radiologist's written report and the MDTs. The report written by the radiologist with their interpretation of the scan will aim to answer the clinical question that accompanies the scan and will be read by the referrer and any other consultant who has a stake in that patient's care. Any unrelated incidental findings will also be reported. For straightforward cases, this will be the only communication between the reporter and referrer; more complex cases are likely to be sent to an MDT. These MDTs will have at least one of every specialist relevant to the pathology present, and cases will be discussed as a group with each participant putting forward their views. It was made clear by participants that MDTs were introduced to help make better-informed decisions and to lift the responsibility of decisions from 1 person. Participant L described that these meetings aim to "make a good decision quickly."

The reports that accompany scans are the key value that radiologists contribute to the point where, for more straightforward cases, a referrer may not look at the images when planning the next step of the patient's care. Participant A said, "for most relatively simple questions, I would just go by the report."

The MDTs are the main point where cases are discussed and decisions are made with the full range of expertise. During these meetings, the radiologist will share relevant images and talk through the salient details with the group of specialists so that each can put forward their opinion. The images are not likely to be viewed for an extended period of time here, as MDTs are generally a high-level discussion, and there will be a lot of cases to get through in minimal time.

A trend across the interviews was the notion of knowing what information the next clinician in a patient's line of care will need in order to do their job, as well as the radiologist sculpting their report and the presentation of information at the MDT toward that. Participant L said, "I do the same MDT every week and have done for 10 years. So we're a bit more experienced [...] so that we know what they want in those specific circumstances." They then went on to talk about reporting scans from other hospitals and said, "If you don't know your referrers you don't know how they like their reports or whether there are specific things on there they want or things like that. So it's better to report scans from your hospital for a number of

reasons.” There is the idea here that knowing, or having an intuition of the next steps of care, will have an impact on how information is portrayed.

Additionally, it is clear that while radiological images are essential to communicating information and making decisions for patients’ care, they play a supporting role and are only the center of attention to the radiologist reporting them. Each step after this, the radiologist distils the information down to the relevant points, chosen based on experience and specifically to answer relevant clinical questions.

Inconsistencies and Personal Preference in Practice

Overview

This theme encapsulates and describes the observed inconsistencies in practice between the range of specialists interviewed, and how much of a role personal preference plays in the details of practice. This is split into 2 subsections: reporting and tools.

Reporting

Across the dataset, particularly in the interviews with radiologists, the subjective nature of image analysis was made clear. Most of the radiologists used the term “interpretation.”Participant L said, “And my interpretation of it, if someone else has reported it, I will change if I don’t agree with it” in the context of reviewing cases before an MDT meeting. This subjectivity around the details of reporting presented itself directly through radiologists referencing it and also through radiologists talking about confirming others’ “opinions.”Participant L said, “when I’m allocated to do attending [...] we do get a lot of telephone calls asking for opinions from scans which have been done at other hospitals.” The data suggested that the uncertainty was greatest between departments or between hospitals. Participant N said, “if one of my colleagues has reported it [...] usually I just look at what they’ve said, because I’m always going to agree.” This suggests that within departments, experiences and expertise are shared and therefore create an isolated unit of consistency.

Radiologists also talked about sculpting their reports for those who were going to read them. The radiologist participants made it clear that in many cases, they know how specific consultants like their reports or that they know what questions such a consultant would have, and therefore, they write their report for them. This implies a level of inconsistency around what content should be in a report, and that efficiencies are gained by working with the same people for an extended period of time and getting to know how they work. Additionally, part of medical knowledge comes from the scenarios that individuals have experienced and the results of reactions to those scenarios. Participant N recalled 1 difference between him and a colleague who has recently retired was “He’s coming at it with far more experience and that will colour his opinions of all the things he’s seen and the things I haven’t seen. Likewise in certain areas I’ve trained for more recently than he has so some of the more modern things I might have done a little bit more of.” This experiential part of medical knowledge will likely lead to inconsistency in how scans are reported, as different reporters will bring different knowledge and experiences.

Tools

The use of different tools between different departments and trusts was immediately apparent, with personal preference playing a key role.

Picture archiving and communications systems (PACSs) are the systems used in hospitals to store, view, and report radiological images. With many vendors available, it is each NHS trust’s decision which to buy into. While PACS implementations will have a common set of functions, different vendors will have subtly different implementations. This leads to trusts choosing a system that is most appropriate to their specific requirements. As such, interoperability, and in particular, image exchange, between trusts becomes an issue.

The use of third-party tools was a clear example of personal preference throughout the interviews. Third-party tools are a department-level decision, and as such, there was considerable variation in the choices made. Participant J said, “we haven’t bought into any of that market [...] because we think at the moment, if you have a one millimeter or less slice contrast-enhanced scan, with our PACS system, you should be able to reconstruct and see sufficiently.” Conversely, participant N had more than 1 third-party tool available to use and described 1 of the third-party tools they use as “fairly ubiquitous in cardiac MRI.”This demonstrates that there is some consistency regarding the tools that are used within specialties, but across specialties, there are differing views toward the built-in tools available in PACS systems.

Throughout the interviews, there was a lot of conversation about 2D versus 3D methods of viewing cross-sectional radiological images, such as CT images. Most participants saw 2D slices as enough. Participant I said, “You scroll through [the 2D slices] using the mouse wheel and I’m building up a picture going through the images. And I have to say that’s more than enough.” Other participants, both radiologists and surgeons, said similar things. 3D techniques were used in specific scenarios, such as looking at the whole surface of a structure, such as the skull, as noted by participant K. Generally, 3D images were used for specific questions, but participants claimed they did not add very much value beyond that.

However, there were situations where 3D techniques were very valuable. Participant H, a thoracic lung surgeon, used a third-party company to reconstruct cross-sectional scans into highly accurate 3D models to be able to plan their operations better. They commended its value, but due to the cost per case, said it cannot be used for every patient; they said, “the frustration is that we can’t have it for every single patient.”It is clear then that traditional 2D techniques are still dominant, but in certain groups, and in certain scenarios, newer 3D techniques are adding value.

XR Maturity for Surgery

It was clear throughout the study that radiologists, surgeons, and other consultants have very different relationships with radiological images. This is unsurprising, but the analysis was an opportunity to delineate these relationships and understand the effect that they have on experiences and requirements of current AR or XR systems.

It was evident that radiologists spend a much more extended period looking at images as they have a much broader question to answer. Radiologists will answer the clinical question that accompanies the set of images, but they will also look at the rest of the pictured area and report "incidental findings" if required. These incidental findings are a key point of value that the radiologists add. Surgeons, on the other hand, will be looking to answer very specific questions that may affect or change the operation they are about to conduct. One radiologist participant summarized this difference as "If you have a brain surgeon they're going to be an expert in looking at things they can operate on [...] But if you showed them something they can't operate on, like a stroke, they're not going to recognise it. The radiologist adds value in looking at all the other things on the scan." An example of this would be a radiologist measuring a key structure pictured in the scan and including this measurement in their report. The surgeon would then take this measurement as information to use when deciding whether or not to operate or when planning how to approach the procedure.

Across all participants, the experience of AR in clinical practice was little to none, and the opinions of current systems were consistent, particularly among the surgeon participants. The view of the current systems on the market indicated that they added very little value, and definitely not enough to overcome the cost of buying into such technology. Participant E, a cardiac surgeon, referred to the systems they had experienced as "perhaps not quite at the gimmick end of the spectrum, moving a little bit away from that, but still there." There was some inconsistency around opinions as to what role AR could play in the future. Some participants could very much see the potential value in specific areas, while others could not see how AR could improve their current capability or practice in any way. Participant H, an orthopedic surgeon, looked into using a Microsoft HoloLens (Microsoft Corp) to guide the placement of implants, while participant I, a general surgeon, said the presentation of scans as 2D slices is "more than enough"to get the information they need to operate successfully. Participant H acknowledged the potential value of AR for thoracic lung surgery but reinforced the importance of correct registration and how this remains an unsolved issue with the current state of the art.

One of the first things AR was suggested for is IGS, and it is one of the applications that could be most valuable [45]. Most of the surgeons spoken to in this study saw some role for AR to aid surgery as being in the future of their fields. IGS has a very wide scope with many different surgical fields and specific interventions that could benefit from AR, and each will have its own requirements. Robotic surgery is an obvious potential application, as the surgeon is already looking at the operating site through a headset of sorts. Participant J, a robotic thoracic surgeon, when asked about the future said, "there's got to be more things that can be fed into your vision during your operation" and commending the potential of guidance as a way to reduce risk to patients they said, "there have been times, don't get me wrong, where I wonder where I am in the chest, and an overlay at that point would be delightful because your fear factor has gone up." This is a demonstration of where AR could provide tangible value in IGS. It may not be all surgical fields that benefit in this way, though; AR may be introduced in another way. Participant E, a cardiac surgeon, struggled to see how AR could help in their field. Given this constraint, AR may be applied in a different way to add value, such as acting as a head-up display with information like the patient's vital statistics or a view of the preoperative scans floating above the body to act as guidance in a different way.

Increased Reliance on Imaging

Overview

An increased reliance on imaging is a known issue in radiology [8] within the NHS and has multiple contributing factors, but this is likely to have ramifications throughout the organization. Across the interviews, the requirement for efficiency was ever-present, particularly with the radiologists, as were the acceptability factors that new technologies have to work through in the medical field.

Efficiency

Already, there are more scans being taken than can be reported by radiologists, and this is likely to only increase [8] as imaging is an essential part of modern practice [46]. AI for reporting radiological images was brought up regularly in the interviews when talking about the future and efficiency in particular. It was nearly unanimous across all the participants who spoke about it that it would have a big impact on radiology reporting and the number of scans that could be reported in a given time. With an increasing demand being placed on radiologists, the backlog of images to be reported will only grow, increasing waiting times for patients and potentially having negative effects on their care. There was, however, disagreement over exactly how AI would be used. Participant I, a general surgeon, said, "in theory you could replace a radiologist with a computer,"and this was shared among a few others. However, the radiologists saw AI, at least in the foreseeable future, as a tool for radiologists rather than a replacement. Participant K, a neuro-radiologist, said, "having worked in radiology for 6 years and now a year into being a consultant, I think it's difficult to ever imagine a world in which AI could do everything that a radiologist does," and participant N, a thoracic radiologist, said, "AI's got to get pretty good before it's able to do that because that requires a lot of higher functioning and thought [...]-It's a tool, and I see it as a tool going forward."

Similarly to the reporting process, as more imaging is used, MDTs will have to discuss it, and therefore, the process of viewing and manipulating images will have to become more efficient. Radiologists attending an MDT will likely have to review many scans that may have been reported by someone else, quickly, as preparation. Participant N said, "you only get a couple of minutes per case to prep the MDT. Because obviously there's quite a lot of cases, so I couldn't realistically re-report every single scan." It is here that the radiologists check that they agree with how the scan has been reported, particularly in uncertain or complex cases.

Acceptability

When looking to the future of medical technology, there were several factors that repeatedly surfaced through the interviews.

The first was phrased well by participant N as “technology inertia,”which captures well the resistant nature of the medical field. They went on to say, “I think it’s [the medical field] less open [to new tech], because of the stakes.”This is compounded by other participants saying things such as “people get used to a way of doing things.”This all suggests that even if a new, better technology is available, it takes a significant investment in time and money to implement it in practice. Consultants do not have the time to retrain on new equipment for a very small gain in performance. Current methods are quick through experience and practice and are therefore preferred to retraining. There is a positive attitude toward new and beneficial technology, as evidenced by participant A who said, “I quite like moving with new ideas where possible.”However, this is inconsistent between consultants and not always reflected in the uptake of new technology.

Where there was mention of resistance to new technology, there was often the mention of how age affected this. Participant N said, “to some extent you do rely on younger colleagues coming through to help you innovate, I guess.”This adds to the line of thought that even though new technology may be an improvement, it takes a push to get through the inertia. Just as younger colleagues help the more established to innovate, we must provide a means by which new technology can be effectively demonstrated in order to overcome this inertia.

Discussion

Principal Findings

Across all of the themes described in the results section above, there were several linkages. Efficiency came up explicitly and implicitly throughout the interviews, and this is reflected in the themes. There is a persistent reference toward the fact that there are more images taken than can be reported and that this workload is likely to increase [8]. In this vein, there is generally a positive view that new technology has value to provide the medical field, but a contradictory view that current tools, systems, and processes are good enough to obtain the results required and to do the job well. The opportunity here is to understand the clinical requirements and issues being faced and suggest how AR could be used to alleviate this pressure. This section takes the above results and presents 3 design implications as an output, which stand as the core contribution of this work. These design implications, presented at the end of the following subsections, are intended as considerations to be made when investigating the development of AR systems within health care.

After the interviews had taken place and the analysis had been completed, one of the participants was approached to join as a coauthor (author 4). Here, author 4 reviewed the presentation of the clinical side of the analysis and provided further clinical context to the design implications that are presented below.

Where to Go With the Current Technological Capability

AR has a great deal of value to offer, but it is an emerging technology [47] and has limitations that need to be taken into account when applying the technology. It is important to acknowledge the capabilities of the state-of-the-art AR technology, as well as its limitations. To have this technology deployed in this sector, there must be proof of the value within the limitations. As discussed previously, AR cannot currently reach acceptable margins for IGS. However, over time, as the technology develops, the technological limitations will dissipate, and applications that demand tight margins, such as IGS, will become more feasible. Once AR can be proven to function within the acceptable margins of IGS, there is huge value to be gained [25]. Many of the surgeons interviewed saw the potential value of AR IGS, and the literature supports this [45]. Before this happens, AR still has value to exploit, and it must be determined where the technology can be used to make a difference in its current form. In this section, we suggest radiology as an initial application for integrating AR.

Two key recurring points in our analysis are important here: the desire for efficiency in the workflows around radiological imaging, particularly from radiologists, and the ways in which images are engaged with at each stage of the workflow. Our analysis suggests that there are 2 important points of communication regarding radiological images: the radiologist’s written report and the MDTs. In both the report and the MDTs, the images are, of course, integral, but the time spent on the images after they have been reported can be minimal. This is, in particular, in situations where there is a relatively simple case and the radiologist knows which consultant will be reading the report. They are therefore able to pre-emptively answer the questions the consultant is likely to ask. This matter of minimal time spent looking at images continues to the surgical planning stage. All the surgeons interviewed said that this was a short task where they were looking to answer specific questions that would impact the feasibility of an operation or how an operation would be performed, not a complete reevaluation of the images.

The requirement for efficiency came up repeatedly, particularly from the radiologists’ point of view, and this is consistent in the literature [8]. As discussed previously, the reporting of scans is going to have to become more efficient as the number of scans taken already exceeds the number of scans that can be reported. This extends to the radiologists’ preparation for MDTs, where each case must be reviewed by the radiologist attending the MDT in advance.

In response to these points, we suggest radiology as a starting point for integrating AR into health care, as we believe that the inherent interaction benefits of AR are well placed to be exploited when viewing 3D images. This could give radiologists a better appreciation of the anatomy in a shorter period of time and help them understand relationships between key structures. It may also be used here to take more accurate, quicker measurements of key structures that could help surgeons be better prepared for interventions. This could be of benefit in terms of efficiency.

In addition to this, radiologists spend a significant amount of time with the scans for each case, much more than any other clinician at any other stage in the workflow. This means that the value of using AR can be maximized, and limitations such as the cost of equipment and the learning curve of using it are limited.

Establishing AR in radiology could then allow some usability, procedural, and technological issues to be researched further as part of this deployment of AR. This could then prepare the technology for future deployment in scenarios where there are currently other limitations. Using this as an opportunity to research AR usability in health care, while adding value to the clinical workflow, would be invaluable, as usability issues are as much of a limiting factor to the implementation of AR as technological issues.

This leads to our first design implication: acknowledging AR technology's limitations and the benefits it can provide, namely the interaction potential, AR should be exploited to help increase the efficiency of radiologists reporting scans. This should be followed by clinical evaluations proving the efficacy of the technology, which may then encourage research into expanding the technology into other disciplines as the technological limitations are mitigated with continued development.

Acknowledging the technology's limitations and working with its advantages will allow value to be added to processes almost immediately. We argue that radiologists are well placed to exploit value from the interactions that existing AR technology affords, likely resulting in increased efficiency; whether that is the whole reporting process or a subset of tasks such as taking measurements.

3D Views Complement 2D Views

Throughout the interviews, 2D versus 3D viewing methods of cross-sectional scans, such as CT and MRI, were a key discussion point. The overwhelming majority were of the opinion that 2D slices of scans in 3 planes were more than enough to gain the information that they required. Some went on further to say that 3D methods lose something over 2D because it is more difficult to look at the internal structures. This was contested in a minority of situations where 3D methods had various specific application areas, such as looking at the surface of the skull and reconstructing lung scans for planning resections. The general consensus was that 3D reconstructions are useful for very specific tasks but add little beyond that.

This suggests, and is intuitive, that the main issue with 3D methods for the participants is the inability to see the same internal structure information that is shown with traditional 2D slices. There was no direct issue with 3D forms; rather, the current 3D viewing methods do not add any value. The opportunity here is to use AR to provide the same information that traditional 2D slices provide while adding value with the third axis. This may enable the radiologist to appreciate the information of the internal structures in the context of the full 3D form in a more intuitive manner. This could also enhance communication and allow a greater shared understanding.

There are examples of using AR in such ways [48], but this interaction has yet to be proven. In order to be accepted by radiologists, the scans shown in 3D in AR must show at least as much information as 2D slices while providing additional value in some other way, such as an enhanced interaction. This value is likely to be in the interaction, as viewing 3D anatomy in 2D images is less intuitive than viewing it in 3D, where further context and relationships may be more visible. The point here is to demonstrate the additional value that AR can provide. This may be difficult, as our analysis suggested that the medical field is quite resistant to change and new technologies. But if it can be demonstrated well and the value translates into better appreciation of structures, quicker turnaround time, or higher throughput, AR will likely become commonplace in radiology offices.

There is clearly big potential in AR IGS, our analysis and the literature [25] show this, but both also show that it is one of the most challenging areas of research. As discussed previously, there are multiple technological issues and usability issues that need to be resolved to unlock this value that are well documented in the literature, with some suggesting that usability considerations of AR are among the most significant potential barriers to the technology's success [19]. A creative, out-of-the-box approach to these usability problems could allow the successful implementation of AR in health care and, therefore, be a source of great value, allowing the benefits that the technology affords to be exploited in a much wider number of scenarios.

Here we argue for the creative implementation of AR, playing to the strengths of the technology and not simply recreating existing capability in a new medium.

As with the example above, using 3D viewing methods has limited use in current practice, and 2D views are dominant. But given the third axis and immersiveness that AR provides, do 3D views provide something that is difficult in 2 dimensions? For example, better appreciation of complex relationships between structures. Or are 3D images easier to interact with, providing an easier or more accurate way to take measurements of structures of interest?

Designers must be explicit about why AR is appropriate for the application and what value it provides while using creative practices in order to realize the full potential of AR. This is the second design implication we suggest: creativity must be used in the implementation of AR; simply recreating existing capability in a new medium should be avoided, and the strengths of AR should be played to in order to add value to the clinical scenario while maintaining prior ability. In the context of 2D versus 3D images, this could mean that the information provided with 2D slices is still available, while also providing additional contextual information with the third dimension.

What Does an "Augmented-Reality-First" World Look Like?

Our analysis suggests that there would be limited value in applying current AR technology individually to surgical planning or for use in MDTs, as current imaging techniques give consultants adequate information to make the decisions necessary in these situations. Furthermore, the images themselves are not used for a very long period for these tasks, and as such, the value gained from viewing the images in AR would have to be great in order to be worth the cost of the equipment and the time taken to put on, boot up, and engage with an AR headset. This is in addition to the initial strain of rewriting procedures around the new technology and the learning curve of engaging with the new medium.

This can be held true for today's "desktop-first" world, where keyboard and mouse are universally dominant. But looking down the road as AR technology develops and its presence increases in daily life, this is likely to change. In this scenario, where an AR headset could be an extension to a desktop environment, the previous limitations (of cost, learning curve, and clinical practice adjustment) are negated, and the cost-benefit ratio of AR in these situations becomes more amenable.

In this "AR-first" world, the use of an AR headset is as embedded in practice as the use of a normal monitor. There is likely to be a set of tasks that clinicians complete that could be improved in some way with AR. Reporting scans, MDTs, and surgical planning could be 3 examples. For these tasks, the headsets would be ready to run alongside, or instead of, the main desktop environment, and as such, the setup and engagement obstacles are averted. AR would be seamlessly integrated into practices, enabling the benefits to be exploited and made the most of. It is this concept of integration that came up repeatedly in different forms throughout the analysis, for example, learning curve, rewriting processes, resistance to new technologies, and efficiency.

Thinking about speculative scenarios such as this, where certain obstacles are put to one side, allows us to highlight other potentially more nuanced concerns and opportunities that should be considered when designing AR applications for this space. It also allows speculative consideration of the breadth of value the technology could bring in isolation, without being overshadowed by current technological or procedural limitations.

The integration of any new technology into clinical practice can be as significant a hurdle as developing the technology itself, with many concerns residing under the umbrella of "integration"; things such as cost, learning curve, and the rewriting of procedures. However, for the AR, what could be gained if the technology is successfully integrated in the right places? Our analysis suggests that AR brings value in its versatility. It will never be at its best if only used for 1 task. The highest value will be attained when many AR-enhanced tasks are considered. If an AR headset were integrated into practice and ready to deploy for several smaller tasks (such as reporting scans, discussing images in MDTs, and viewing images for surgical planning), much more value would likely be gained relative to implementing just one of those examples.

The first hurdle of successfully integrating AR into 1 point in a workflow and proving value for this one task will likely result in the technology cascading into surrounding tasks, slowly reaching toward maximizing the cost-benefit ratio.

Our analysis suggests 2 main factors would have to be proven to enable an "AR-first" environment. First, is the cost-benefit ratio of the technology. It must be demonstrated that the number of tasks AR could be used for and the benefit that it provides in each of them is worth the cost of buying into the technology. Second, the technology must be integrated into practices well enough to the point where putting on and starting up the headset is not an obstruction to the work being done. This will be a significant challenge as it requires the rewriting of some practices and, therefore, a learning curve when using the systems for the first time. It also requires more targeted human-centered HCI research as opposed to a sole focus on the development of AR technology. Targeted HCI research could map this space more effectively, solving some usability issues and laying the groundwork for more advanced AR technology to stand on.

This leads to the final design implication: AR brings value through its versatility. To obtain the most out of this versatility, it must be considered how AR tools integrate with existing workflows and how they will be used in order to create a seamless transition toward wider uptake of the technology. The technology should be integrated in such a way that negative disruption to existing workflows is avoided and maximum value can be gained from multiple workflows.

Future Work

These design implications aim to help direct and inform future research, while also aiding in decision-making when developing AR applications in this space. Future work will develop these design implications further and test their feasibility by developing a case study application. This case study will conduct further user research and then incorporate the outcomes of this with these design considerations into a prototype. This prototype will then be evaluated by users against the design implications.

This work could also be expanded by focusing on medical education and training. We chose to focus on the clinical radiological applications of AR for this study to contain the scope and focus the design implications. However, participants mentioned educational and training applications, and there is literature supporting their development. Future work could be done to expand or develop these design implications in this space.

Limitations

Our qualitative analysis aims to provide a representative insight into the views and opinions of hospital-based consultants in the United Kingdom along with their views on AR and the role it could play in radiological imaging. However, we must acknowledge the limitations of both the methodology and the dataset.

Our participants were hospital-based consultants, largely from the North East of England, with a few from the North West and South. We successfully recruited a range of participants with a range of specialisms to provide a variety of views and differing contexts, which adds strength and breadth to this work. However, a potential shortcoming of this participant pool was our ability to only recruit men. Where possible, we took appropriate steps to try and recruit women, but in part due to this being a very male-dominated field [49], we were unable to. This will restrict the gender diversity of the perspectives presented, but it reflects the wider demographic trend in some specialties. Future work should aim for a more diverse participant pool.

Our study was limited to the United Kingdom, which we acknowledge may limit the generalizability to wider audiences. However, this limitation is commensurate with the scope of this work.

We also focused heavily on AR for radiological imaging with little mention of AR for education or training. That is not to say that AR should not be applied to these areas, and it was brought up by participants in multiple interviews. However, for this study, we chose to focus on AR for radiological imaging in order to focus on the design implications.

Conclusions

In this paper, we have presented the results of a thematic analysis of interviews with hospital-based consultants in order to investigate the role AR could play in radiological imaging. We contribute 3 design implications for AR systems within radiological imaging workflows based on the results of our qualitative analysis and frame them in the context of the HCI and medical fields.

The first design implication outlines the desire for efficiency. AR has the potential to provide enhanced interactions, which could allow for a better appreciation of the anatomy and quicker measurements. Radiologists are well placed to exploit this value as a tool to improve efficiency because being able to view and interpret images quickly would allow them to have a higher throughput. Second, we suggest that AR tools need to be built in such a way that no capability available with existing 2D desktop workflows is lost either by using AR to complement existing 2D workflows or by integrating the 2D capability into AR. Finally, AR tools need to integrate and be interoperable with existing radiology systems to minimize disruption to existing workflows, for example, ensuring compatibility with PACS. The value of AR could be exploited across health care organizations if the technology is integrated well, and we speculate on the impact of what an “AR-first” world may look like and how clinical practices may change were this to happen.

This work also adds to the body of literature acknowledging active surgeons’ opinions toward the potential value of AR IGS and motivates areas of future research into AR’s place around radiological images.

Conflicts of Interest

None declared.

References

1. Kelly PJ, Alker GJ, Goerss S. Computer-assisted stereotactic laser microsurgery for the treatment of intracranial neoplasms. Neurosurgery 1982 Mar 1;10(3):324-331. [doi: 10.1227/00006123-198203000-00005]
2. Rosenberg L. The use of virtual fixtures as perceptual overlays to enhance operator performance in remote environments. : Armstrong Laboratory; 1992 URL: https://apps.dtic.mil/sti/html/tr/ADA292450/ [accessed 2025-08-01]
3. Feng Z, Been LDH, Mark B. Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR. Presented at: 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality; Sep 15-18, 2008; Cambridge. [doi: 10.1109/ISMAR.2008.4637362]
4. Dhar P, Rocks T, Samarasinghe RM, Stephenson G, Smith C. Augmented reality in medical education: students’ experiences and learning outcomes. Med Educ Online 2021 Dec;26(1):1953953. [doi: 10.1080/10872981.2021.1953953] [Medline: 34259122]
5. Vinci C, Brandon KO, Kleinjan M, Brandon TH. The clinical potential of augmented reality. Clin Psychol: Sci Pract 2020;27(3):e12357. [doi: 10.1111/cpsp.12357]
6. Eckert M, Volmerg JS, Friedrich CM. Augmented reality in medicine: systematic and bibliographic review. JMIR mHealth uHealth 2019 Apr 26;7(4):e10967. [doi: 10.2196/10967] [Medline: 31025950]
7. Barcali E, Iadanza E, Manetti L, Francia P, Nardi C, Bocchi L. Augmented reality in surgery: a scoping review. Appl Sci (Basel) 2022;12(14):6890. [doi: 10.3390/app12146890]
8. Clinical oncology census reports. Royal College of Radiologists. 2023. URL: https://www.rcr.ac.uk/news-policy/policy-reports-initiatives/clinical-oncology-census-reports [accessed 2025-08-01]
9. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport, Exercise and Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
10. Braun V, Clarke V. Thematic Analysis: A Practical Guide: SAGE Publications Ltd; 2021. URL: https://uk.sagepub.com/en-gb/eur/thematic-analysis/book248481 [accessed 2025-08-01]
11. Caudell TP, Mizell DW. Augmented reality: an application of heads-up display technology to manual manufacturing processes. Presented at: Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences; Jan 7-10, 1992; Kauai, HI. [doi: 10.1109/HICSS.1992.183317]
12. McCarthy CJ, Uppot RN. Advances in virtual and augmented reality—exploring the role in health-care education. J Radiol Nurs 2019 Jun;38(2):104-105. [doi: 10.1016/j.jradnu.2019.01.008]
13. Domingues GC, Vieira V, Yoshida L, et al. What if video see-through in hmds changes how accurately we perform tasks? Presented at: Proceedings of the 26th Symposium on Virtual and Augmented Reality; Sep 30 to Oct 3, 2024; Manaus, Brazil. [doi: 10.1145/3691573.3691599]
14. Yeung AWK, Tosevska A, Klager E, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. J Med Internet Res 2021 Feb 10;23(2):e25499. [doi: 10.2196/25499] [Medline: 33565986]
15. Sun P, Zhao Y, Men J, et al. Application of virtual and augmented reality technology in hip surgery: systematic review. J Med Internet Res 2023 Mar 10;25(1):e37599. [doi: 10.2196/37599] [Medline: 36651587]

16. Westermeier F, Brübach L, Wienrich C, Latoschik ME. A virtualized augmented reality simulation for exploring perceptual incongruencies. Presented at: Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology; Oct 9-11, 2023; Christchurch, New Zealand. [doi: 10.1145/3611659.3617227]
17. Lystbæk MN, Pfeuffer K, Langlotz T, Grønbæk JES, Gellersen H. Spatial gaze markers: supporting effective task switching in augmented reality. Presented at: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, HI. [doi: 10.1145/3613904.3642811]
18. Prinz LM, Mathew T. Support lines and grids for depth ordering in indoor augmented reality using optical see-through head-mounted displays. Presented at: Proceedings of the 2024 ACM Symposium on Spatial User Interaction; Oct 7-8, 2024; Trier, Germany. [doi: 10.1145/3677386.3682097]
19. Eddie PJ. The challenge of augmented reality in surgery. In: Digital Surgery: Springer International Publishing; 2021:121-135. [doi: 10.1007/978-3-030-49100-0_10]
20. Malhotra S, Halabi O, Dakua SP, Padhan J, Paul S, Palliyali W. Augmented reality in surgical navigation: a review of evaluation and validation metrics. Appl Sci (Basel) 2023;13(3):1629. [doi: 10.3390/app13031629]
21. Ferrari V, Cattari N, Fontana U, Cutolo F. Parallax free registration for augmented reality optical see-through displays in the peripersonal space. IEEE Trans Vis Comput Graph 2022 Mar;28(3):1608-1618. [doi: 10.1109/TVCG.2020.3021534] [Medline: 32881688]
22. Gabbard JL, Mehra DG, Swan JE. Effects of AR display context switching and focal distance switching on human performance. IEEE Trans Vis Comput Graph 2019 Jun;25(6):2228-2241. [doi: 10.1109/TVCG.2018.2832633] [Medline: 29994003]
23. Erkelens IM, MacKenzie KJ. 19 - 2: vergence - accommodation conflicts in augmented reality: impacts on perceived image quality. Symp Digest of Tech Papers 2020 Aug;51(1):265-268. [doi: 10.1002/sdtp.13855]
24. Condino S, Carbone M, Piazza R, Ferrari M, Ferrari V. Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks. IEEE Trans Biomed Eng 2020 Feb;67(2):411-419. [doi: 10.1109/TBME.2019.2914517] [Medline: 31059421]
25. Birlo M, Edwards PJE, Clarkson M, Stoyanov D. Utility of optical see-through head mounted displays in augmented reality-assisted surgery: a systematic review. Med Image Anal 2022 Apr;77:102361. [doi: 10.1016/j.media.2022.102361] [Medline: 35168103]
26. Eves J, Sudarsanam A, Shalhoub J, Amiras D. Augmented reality in vascular and endovascular surgery: scoping review. JMIR Serious Games 2022 Sep 23;10(3):e34501. [doi: 10.2196/34501] [Medline: 36149736]
27. Fuchs H, State A, Pisano ED, et al. Towards performing ultrasound-guided needle biopsies from within a head-mounted display. In: Goos G, Hartmanis J, Höhne KH, Kikinis R, editors. Visualization in Biomedical Computing: Springer; 1996. [doi: 10.1007/BFb0047002]
28. Brunet JN, Mendizabal A, Petit A, Golse N, Vibert E, Cotin S. Physics-based deep neural network for augmented reality during liver surgery. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Lecture Notes in Computer Science: Springer; 2019:591-600. [doi: 10.1007/978-3-030-32254-0_16]
29. Bertolo R, Hung A, Porpiglia F, Bove P, Schleicher M, Dasgupta P. Systematic review of augmented reality in urological interventions: the evidences of an impact on surgical outcomes are yet to come. World J Urol 2020 Sep;38(9):2167-2176. [doi: 10.1007/s00345-019-02711-z] [Medline: 30826888]
30. Dilley JWR, Hughes-Hallett A, Pratt PJ, et al. Perfect registration leads to imperfect performance: a randomized trial of multimodal intraoperative image guidance. Ann Surg 2019 Feb;269(2):236-242. [doi: 10.1097/SLA.0000000000002793] [Medline: 29727330]
31. Palumbo A. Microsoft HoloLens 2 in medical and healthcare context: state of the art and future prospects. Sensors (Basel) 2022 Oct 11;22(20):7709. [doi: 10.3390/s22207709] [Medline: 36298059]
32. Parsons D, MacCallum K. Current perspectives on augmented reality in medical education: applications, affordances and limitations. Adv Med Educ Pract 2021;12:77-91. [doi: 10.2147/AMEP.S249891] [Medline: 33500677]
33. Rangarajan K, Davis H, Pucher PH. Systematic review of virtual haptics in surgical simulation: a valid educational tool? J Surg Educ 2020;77(2):337-347. [doi: 10.1016/j.jsurg.2019.09.006] [Medline: 31564519]
34. Dallas-Orr D, Penev Y, Schultz R, Courtier J. Comparing computed tomography-derived augmented reality holograms to a standard picture archiving and communication systems viewer for presurgical planning: feasibility study. JMIR Perioper Med 2020 Sep 24;3(2):e18367. [doi: 10.2196/18367] [Medline: 33393933]
35. Izard SG, Méndez JAJ, Palomera PR, García-Peñalvo FJ. Applications of virtual and augmented reality in biomedical imaging. J Med Syst 2019 Mar 14;43(4):102. [doi: 10.1007/s10916-019-1239-z] [Medline: 30874965]
36. Douglas DB, Wilke CA, Gibson JD, Boone JM, Wintermark M. Augmented reality: advances in diagnostic imaging. MTI 2017;1(4):29. [doi: 10.3390/mti1040029]
37. Pelargos PE, Nagasawa DT, Lagman C, et al. Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery. J Clin Neurosci 2017 Jan;35:1-4. [doi: 10.1016/j.jocn.2016.09.002] [Medline: 28137372]
38. Elsayed M, Kadom N, Ghobadi C, et al. Virtual and augmented reality: potential applications in radiology. Acta Radiol 2020 Sep;61(9):1258-1265. [doi: 10.1177/0284185119897362] [Medline: 31928346]

39. Trestioreanu L, Glauner P, Meira JA, Gindt M, State R. Using augmented re- ality and machine learning in radiology. In: Glauner P, Plugmann P, editors. Innovative Tech- Nologies for Market Leadership: Investing in the Future, Future of Business and Finance: Springer International Publishing; 2020:89-106. [doi: 10.1007/978-3-030-41309-5_8]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Fine M. Disruptive Voices: The Possibilities of Feminist Research: University of Michigan Press; 2021.
42. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern‐based qualitative analytic approaches. Couns and Psychother Res 2021 Mar;21(1):37-47. [doi: 10.1002/capr.12360]
43. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches, 4th edition: SAGE Publications; 2016.
44. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2021 Jul 3;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
45. Glas HH, Kraeima J, van Ooijen PMA, Spijkervet FKL, Yu L, Witjes MJH. Augmented reality visualization for image-guided surgery: a validation study using a three-dimensional printed phantom. J Oral Maxillofac Surg 2021 Sep;79(9):1943. [doi: 10.1016/j.joms.2021.04.001] [Medline: 34033801]
46. Łoginoff J, Augustynowicz K, Świąder K, et al. Advancements in radiology and diagnostic imaging. J Educ Health Sport 2023;33(1):45-51. [doi: 10.12775/JEHS.2023.33.01.005]
47. Xiong J, Hsiang EL, He Z, Zhan T, Wu ST. Augmented reality and virtual reality displays: emerging technologies and future perspectives. Light Sci Appl 2021 Oct 25;10(1):216. [doi: 10.1038/s41377-021-00658-8] [Medline: 34697292]
48. apoQlar. 2024. URL: https://apoqlar.com [accessed 2025-08-01]
49. Barnes KL, McGuire L, Dunivan G, Sussman AL, Kee RM. Gender bias experiences of female surgical trainees. J Surg Educ 2019;76(6):e1-e14. [doi: 10.1016/j.jsurg.2019.07.024]

Abbreviations

AR: augmented reality
CT: computed tomography
HCI: Human-Computer Interaction
HMD: head-mounted display
IGS: image-guided surgery
MDT: multidisciplinary team
MRI: magnetic resonance imaging
NHS: National Health Service
PACS: picture archiving and communications system
XR: extended reality

Edited by L Riedemann; submitted 14.11.24; peer-reviewed by H Tanabe, J Zhang, M Agbede, R Uppot; revised version received 13.06.25; accepted 11.07.25; published 14.08.25.

Please cite as:
Hobbs J, Bull C, Claisse C, Elameer M, Davison R
Role of Augmented Reality in Tertiary Care: Qualitative Investigation Using Thematic Analysis
JMIR XR Spatial Comput 2025;2:e68810
URL: https://xr.jmir.org/2025/1/e68810
doi:10.2196/68810

At-Home Virtual Reality Intervention for Patients With Chronic Musculoskeletal Pain: Single-Case Experimental Design Study

Syl Slatman[1,2], MSc; Lieke Heesink[3], PhD; Reinoud Achterkamp[4], PhD; José Broeks[4], MSc; Nelson Monteiro de Oliveira[5], MD; Remko ter Riet[5], MSc; Marjolein Stegeman[4], MSc; Monique Tabak[1], Prof Dr, Ir.

[1]Biomedical Signals and Systems Group, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Drienerlolaan 5, Enschede, Netherlands

[2]Musculoskeletal Rehabilitation Research Group, School for Allied Health, HAN University of Applied Sciences, Nijmegen, Netherlands

[3]Health Technology and Services Research, Faculty of Behavioural, Management and Social Sciences, University of Twente, Hallenweg 5, Enschede, Netherlands

[4]Department of Chronic Pain, Roessingh Rehabilitation Center, Enschede, Netherlands

[5]Nocepta, Hengelo, Netherlands

Corresponding Author:
Lieke Heesink, PhD
Health Technology and Services Research, Faculty of Behavioural, Management and Social Sciences, University of Twente, Hallenweg 5, Enschede, Netherlands

Abstract

Background: Virtual reality (VR) could possibly alleviate complaints related to chronic musculoskeletal pain (CMP); however, little is known about how it affects pain-related variables on an individual level and how patients experience this intervention.

Objective: This study aimed to gain detailed insight into the influence of an at-home VR intervention for pain education and management on pain-related variables, and to explore its feasibility and general experience.

Methods: The study applied a single-case experimental design in which an at-home VR intervention was used for 4 weeks by patients with CMP who were on a waiting list for regular pain treatment. Outcome measures included pain-related variables, functioning, and objectively measured outcomes (ie, stress, sleep, and steps). Outcomes were analyzed using data visualization (based on line plots) and statistical methods (ie, Tau-U and reliable change index) on an individual and group level. In addition, a focus group was conducted to assess feasibility and general experience to substantiate findings from the single-case experimental design study. This focus group was analyzed using inductive thematic analysis.

Results: A total of 7 participants (female: n=6, 86%) with a median age of 45 (range 31 - 61) years participated in this study. A dataset with 42 measurement moments was collected with a median of 280 (range 241 - 315) data points per participant. No statistically significant or clinically relevant differences between the intervention and no-intervention phases were found. Results of the visual analysis of the diary data showed that patients responded differently to the intervention. Results of the focus group with 3 participants showed that the VR intervention was perceived as a feasible and valued additional intervention.

Conclusions: Although patients expressed a positive perspective on this VR intervention, it did not seem to influence pain-related outcomes. Individual patients responded differently to the intervention, which implies that this intervention might not be suitable for all patients. Future studies should examine which CMP patients VR is effective for and explore its working mechanisms. In addition, future larger trials should be conducted to complement this study's findings on the effectiveness of this intervention for patients with CMP and whether VR prevents deterioration on the waiting list compared with a control group.

(JMIR XR Spatial Comput 2025;2:e58784) doi:10.2196/58784

KEYWORDS

virtual reality; VR; chronic musculoskeletal pain; CMP; single-case experimental design; SCED; user experience; self-management; musculoskeletal pain

Introduction

Chronic musculoskeletal pain (CMP), defined as pain lasting longer than 3 months, is a major problem and prevalent in approximately 20% of adults [1,2]. CMP is associated with a decrease in quality of life and mental health problems [3,4], next to the significant financial and societal burden [1]. Unfortunately, the effectiveness of biomedical treatment options for CMP does not seem to be very promising [5], since CMP usually is a complex problem with an interplay of biological, psychological, and social factors [6].

Given the complexity of CMP, treatment should use a holistic approach in accordance with the biopsychosocial model [5] and neuromatrix theory [7]. Unfortunately, most more complex, holistic interventions for CMP have a waiting list period, which could have a deteriorating effect on patients with CMP [8]. Therefore, it might be sensible to already start treatment during this waiting list period. Virtual reality (VR) is a novel, therapeutic technology that is suitable for stand-alone, at-home treatment [9]. VR is defined as "a collection of technologies that allow people to interact efficiently with 3D computerized databases in real time using their natural senses and skills" [10].

Even though VR for CMP seems promising, much is still unknown about its underlying mechanisms (eg, distraction or skills-building) [11] and influences on an individual level, as previous studies applied a nomothetic approach [9]. Since the principles underlying VR for CMP remain a black box [12], an idiographic approach is warranted for a complex condition like CMP to gain insight into the influence of VR on individual outcomes [13]. A single-case experimental design (SCED) study could increase understanding of the individual experience [14]. SCED studies apply detailed assessment at numerous timepoints [15] and have benefits over other designs, including patients serving as their own control and being especially suitable for heterogeneous samples, like CMP patients with a variety of conditions [16]. A recent SCED study on VR for chronic low back pain (CLBP) found that VR has the potential to reduce CMP-related complaints, possibly through a combination of distraction and modification of attitudes and beliefs [17]. We expect that this VR intervention is suitable not only for patients with CLBP but also for patients with other CMP conditions. In addition, we hypothesize that VR might influence other outcome measures like pain acceptance and interference, functioning, and objectively measured outcomes.

Therefore, the aim of our study was to (1) explore whether and how a VR intervention has an influence on pain-related variables on an individual level and (2) explore the feasibility and general experience of the VR intervention. To do so, patients with CMP received a pain education and management VR intervention at home while they were on a waiting list to receive pain treatment.

Methods

Design

This mixed methods study consisted of 2 parts. The first part of the study applied a nonconcurrent single-case experimental ABA-design on at-home, VR intervention for patients with primary or secondary CMP who were on a waiting list to receive regular pain treatment. Phases A1 and A2 (no intervention) were 1 week before and 1 week after the VR intervention, fulfilling the criterion for a sufficient baseline in single-case designs [18]. Phase B (VR intervention) lasted a total of 4 weeks. To report and conduct the study, the Single-Case Reporting Guideline in Behavioural Interventions (SCRIBE) was used [19], more details in Multimedia Appendix 1. The second part of this study consisted of 1 focus group with patients with CMP who received the intervention. The aim of this focus group was to gain more insight into the general experience and feasibility (including acceptability and practicality, which includes participants' satisfaction and ability to use a new intervention [20]) of the VR intervention and substantiate findings from the SCED study. This part of the study was reported and conducted according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) reporting guidelines [21], more details in Multimedia Appendix 2. Recruitment and completion of the study procedures was from February 2023 to April 2023.

Ethical Considerations

The medical ethics committee of Radboudumc provided a non-WMO (medical research involving human subjects act) waiver (2022‐15829) to conduct this study. The ethics committee of the University of Twente approved this study (RP 2022‐174), as well as local ethics committees of the participating health care organizations. Participants gave written informed consent before any study procedures and received €50 (US $52) for participation in this study after finishing all procedures. All participant data was pseudonymized.

Participants

Participants were recruited from 4 secondary care organizations in the Netherlands (ie, Roessingh Centrum voor Revalidatie, Roessingh Pijnrevalidatie, ZGT Nocepta, and Deventer hospital). Patients were deemed eligible for participation if they (1) were aged 18 years or older, (2) had primary or secondary CMP, (3) finished first-line treatment, (4) were open to treatment with biopsychosocial elements, and (5) were willing and able to comply with the study protocol. Patients were excluded if they (1) were not capable of finishing the intervention due to physical (eg, face wounds, severe visual impairment), mental (eg, severe sensitivity to stimuli), or practical problems (eg, insufficient tech literacy); and (2) had no comprehension of the Dutch language.

Intervention

In this study, the Conformité Européenne (CE)–certified VR intervention Reducept was used as a daily at-home intervention for 10 to 30 minutes per day for 4 weeks, thereby following the intervention protocol dosage from the intervention provider. Besides pain neuroscience education (PNE), the VR intervention incorporates elements of several psychological therapies into 1 application: hypnotherapy, mindfulness, acceptance and commitment therapy (ACT), and cognitive behavioral therapy (CBT). The intervention was described in more detail in previous studies [9,22,23]. The Pico G2 4K (Bytedance) head-mounted display (HMD) was used in this study to provide the immersive VR intervention.

Procedure

Patients visited one of the participating centers of this study for their pain treatment. After their intake, but before starting their secondary care treatment (either [non]invasive pain treatment or interdisciplinary pain rehabilitation), patients were screened by their health care professional for possible participation in the study. Patients were given the opportunity to participate in our study or wait for their treatment on the waiting list without receiving any other treatment. In addition, participants were made clear that participating in this study would not have any influence on the pain treatment they were on a waiting list for. If a patient was deemed eligible, he or she was contacted by

their health care professional, who gave a brief explanation about the study and asked for permission to forward the patient's contact details to the researcher (through a fully secured app: Siilo). Next, the researcher contacted the patient by phone and gave more detailed information about the study and asked the patient to contemplate participating in the study. The patient enrolled in the study by signing the informed consent and received the first questionnaires (T0), the Garmin Forerunner 255 wearable, and the VR headset. The wearable and VR headset were provided by the researcher and used by participants for the duration of the study procedures. In the first week, a detailed baseline was obtained by asking patients to use the wearable and fill in the diary and weekly questionnaires, without receiving the intervention (phase A1). After this phase, participants carried out the intervention at home for four weeks (phase B). Next, patients waited a week (phase A2) before receiving the pain treatment he or she was on the waiting list for. After phase A2 and during the period patients received the pain treatment they were on a waiting list for, patients returned the used equipment (ie, VR headset and wearable) and were invited to the online focus group, using Microsoft Teams, about the feasibility and general experience of the intervention. The focus group was conducted by 2 researchers (SS and LH), assisted by a research student assistant. Both SS and LH attended various courses on and have previous experience with qualitative research. Given this experience, there may have been preconceived notions regarding VR for CMP. We aimed to reduce potential biases by fostering open discussions and critical reflections throughout data collection and analysis. None of the participants had previous relationships with any of the researchers conducting and analyzing the focus group. The topic list used for this focus group is added in Multimedia Appendix 3.

Outcomes

The outcome measures are shown in Table 1. The TIIM app (University of Twente, Enschede, the Netherlands) was used to collect demographic information, diary measures, and weekly questionnaires.

Table . Overview of outcome measurements.

	Pre	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Post
Patient characteristics	✓							
Diary measures		✓	✓	✓	✓	✓	✓	
Weekly questionnaires		✓	✓	✓	✓	✓	✓	
Wearable data		✓	✓	✓	✓	✓	✓	
VR[a] parameters			✓	✓	✓	✓		
Feasibility								✓

[a]VR: virtual reality.

Diary Measures

The daily diary questions consisted of 4 questions, based on the IMMPACT (Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials) recommendations for chronic pain clinical trials [24]: pain intensity (ie, what score would you give your pain today?), pain interference (ie, how burdensome was your pain today?), physical functioning (ie, to what extent did your pain restrict you in doing daily activities today?), and emotional functioning (ie, how was your mood today?). All questions were scored on a 0 (lowest) to 10 (highest) scale. A recent study showed that daily measures of pain and pain-related variables are both valid and reliable [25].

Weekly Questionnaires

Every week, participants were asked to answer 3 questionnaires to measure pain self-efficacy (Pain Self-Efficacy Questionnaire [PSEQ]) [26], pain acceptance (Chronic Pain Acceptance Questionnaire [CPAQ]) [27], and pain coping (Pain Coping Inventory [PCI]) [28]. These questionnaires were the Dutch translation of the original questionnaires, and all were shown to have adequate reliability and validity [29-31].

Wearable Outcomes

The following outcomes were measured using the wearable: physical activity (ie, daily steps), sleep quality, and stress. Daily sleep quality was scored from 0 (worst sleep quality) to 100 (best sleep quality) based on multiple factors, including sleep duration, stress score during sleep, and restlessness. Daily stress was measured using Garmin's stress level from 0 (lowest stress level) to 100 (highest stress level), which is based on the participant's heart rate variability (HRV). More information about the construction of sleep quality and stress as outcome measures in this study can be found in the Garmin manual [32].

Other Outcomes

The following patient characteristics were asked at baseline: age, gender, duration of CMP, comorbidities, pain location, pain medication use, expectation of intervention, occupational situation, education level (based on [33]), and experience with VR for treatment and entertainment.

VR-related parameters that were monitored included usage and module of the VR intervention.

The feasibility of the intervention was explored using usability data (ie, number of minutes used per day) and a semistructured postintervention focus group with patients who received the intervention.

Statistical Analysis

The results of the SCED study were examined using a combination of statistical and visual analyses [34,35]. Phase A1 of each individual participant was observed to determine a stable personal control to note any revealing alterations for the outcome variables measured in phase B. Both within-phase and between-phase analyses were performed and checked for patterns within participants. To determine changes in outcome variables in SCED studies, it is recommended to use the following factors to interpret the data: (1) raw data, (2) central tendency, (3) trend, (4) variability, (5) point of change, and (6) overlap region [15]. All visual plots were constructed using the Shiny SCDA web application [36,37]. Besides this visual analysis, outcomes of the diary questions and wearable data were statistically analyzed using the Tau-U nonoverlap method [38], using a web-based calculator [39]. Effect sizes for Tau-U were interpreted as small (0-.65), medium (.66-.92), or large (>.92) [38]. To gain insight into the relationship between pain-related variables during the intervention, outcomes of the weekly questionnaires were compared on an individual level using the Reliable Change Index (RCI). The RCI was calculated using the pretreatment and posttreatment scores and was considered reliable at 1.96 or more [40]. Clinically important differences in pain intensity were examined between pre- and postintervention, in which a reduction of ≥30% or 2 points was considered clinically important [41]. The recording of the focus group, which had a duration of 50 minutes, was transcribed using Amberscript. This transcript was analyzed using inductive thematic analysis with Atlas.ti (version 24), based on the 6 steps proposed by Braun and Clarke [42]: (1) (re-)read transcript to familiarize with the data, (2) generate initial codes, (3) combine codes into themes, (4) review themes, (5) define themes, and (6) report findings. These steps were completed by 2 researchers (SS and LH) and discussed until consensus was reached. Finally, all authors agreed on the final themes and results identified during this process.

Results

Patient Characteristics

A total of 9 participants enrolled in this study, of which 7 completed the study (Table 2). In addition, 1 participant stopped due to being too busy and 1 participant completed <50% of the questionnaires and was therefore excluded from the analysis. The 7 participants who were included in the analysis provided a median of 280 (range 241 - 315) data points per participant. None of the participants had previous experience with VR. No adverse events were reported by any of the participants from using the VR intervention.

Table . Demographics of participants (n=7).

Participant	Age (years)	Gender	Highest level of education	Occupational situation	Pain duration (years)	Pain location	Medication use	Expectancy[a]
1	31	Woman	Higher	Part-time	1	Foot, ankle	Yes	6
2	55	Man	Lower	Full-time	17	Legs, hands	Yes	5
3	45	Woman	Middle	Part-time	5	Wrist, shoulder, back	Yes	4
4	31	Woman	Middle	Unemployed	7	Generalized	No	6
5	61	Woman	Lower	Part-time	30	Back, hip	Yes	6
6	52	Woman	Higher	Full-time	3	Back, shoulders, neck	Yes	5
7	37	Woman	Higher	Part-time	4.5	Back, pelvic	Yes	6

[a]Scored from 0 (lowest expectancy) to 10 (highest expectancy).

Visual Analysis

Results of the visual analysis of the diary data showed that patients responded differently to the intervention, as discussed below per outcome variable. The results of the 4 diary outcome measures are presented in Figures 1 and 2 and Multimedia Appendix 4, in which the phases A1 (day 1 - 7, no intervention), B (day 8 - 35, intervention), and A2 (day 36 - 42, no intervention) are presented on the x-axis and scores from 0 (lowest) to 10 (highest) are presented on the y-axis.

Figure 1. Visual analysis of diary data on pain intensity (see clearer version in Multimedia Appendix 5).

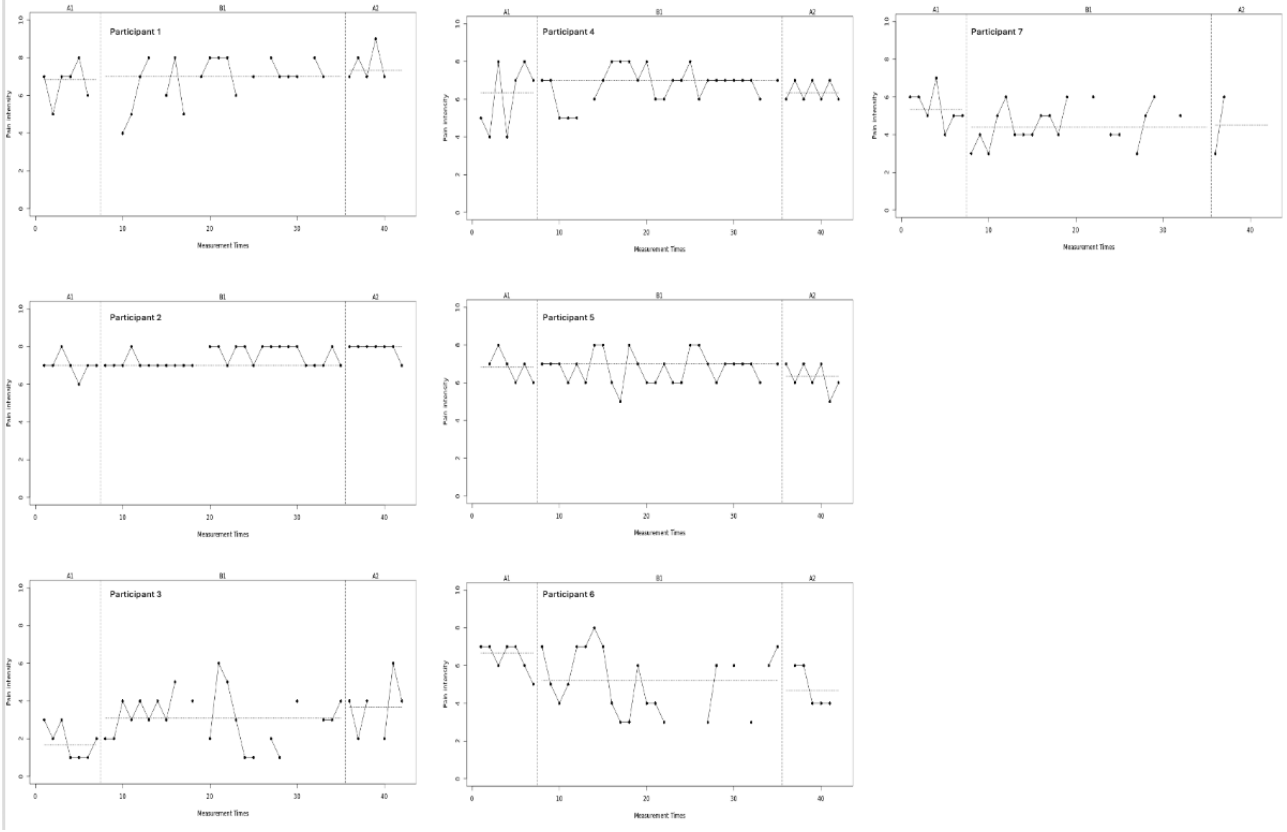
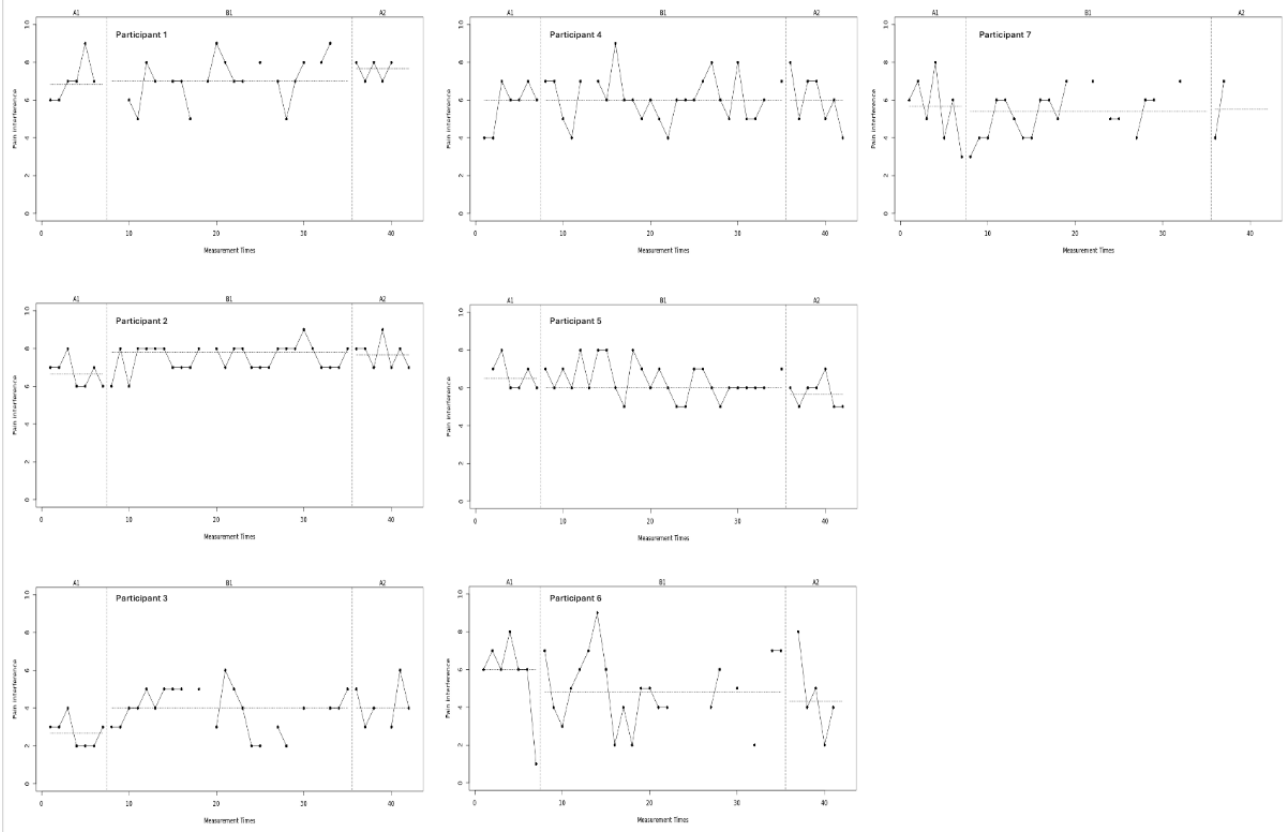


Figure 2. Visual analysis of diary data on pain interference (see clearer version in Multimedia Appendix 6).



Pain intensity scores (Figure 1) remained relatively consistent through phase A1, B, and A2. However, some participants seem to report somewhat lower scores during phase B compared with phase A1 (eg, participant 6 from mean phase A1 6.4, SD 0.8, to mean phase B 5.1, SD 1.7), while others report higher scores (eg, participant 3 from mean phase A1 1.9, SD 0.9 to mean

phase B 3.3, SD 1.4). Furthermore, it is notable that most participants reported substantial variability within proximate measurement moments.

Analysis of the pain interference outcome (Figure 2) showed that patients reported fairly stable scores on central tendency. Some participants showed minor improvement between phases (eg, participant 2 from mean phase A1 6.7, SD 0.8, to mean phase B 7.5, SD 0.7), while others showed some deterioration (eg, participant 5 from mean phase B 6.4, SD 0.9, to mean phase A2 5.7, SD 0.8). In addition, it should be noted that pain interference scores show much likeness to pain intensity scores.

Results on physical functioning (Multimedia Appendix 4) showed that central tendency does not seem to alter too much between phases, similar to the results on pain intensity and pain interference scores. Variability within patients seems to be similar to previously reported outcome measures as well, except for participant 3 who shows large variability within proximate measurement times (eg, day 23: 2; day 24: 10; day 25: 2).

Finally, emotional functioning scores (Multimedia Appendix 4) were relatively high in most participants (mean 7.1, SD 1.5, compared with mean pain intensity 5.9, SD 1.8, pain interference 5.9, SD 1.8, and physical functioning 5.4, SD 1.7). Trend between phases seemed to be improving for some participants (eg, phase A1 of participant 7), while the opposite occurred in other participants (eg, phase A2 of participant 4). Variability seemed to be lower compared with previously discussed outcome measures in most participants.

Statistical Analysis

Analysis of the daily diary and wearable data using Tau-U, as shown in Table 3, showed no statistically significant difference in any of the outcome measures. In addition, no clinically important reductions in pain intensity (ie, reduction of pain intensity score of ≥30% or ≥2 points) were found. Results of the statistical analysis of the weekly questionnaires using the RCI (Table 4) showed no reliable change on any of the questionnaires for any of the participants. More detailed information about the results of the wearable data and weekly questionnaires can be found in respectively Multimedia Appendix 7 (individual scores on steps, stress, and sleep) and Multimedia Appendix 8 (Group scores on weekly questionnaires). Median VR use was 37.5 minutes per week (range 7.8 - 78.4).

Table . Statistical analysis of diary and wearable data.

	Tau-U	95% CI	*P* value
Pain intensity	−0.011	−0.16 to 0.14	.88
Pain interference	−0.013	−0.16 to 0.13	.87
Physical functioning	−0.091	−0.24 to 0.06	.23
Emotional functioning	−0.021	−0.17 to 0.13	.78
Steps	0.013	−0.14 to 0.17	.87
Stress	−0.075	−0.23 to 0.09	.36
Sleep	0.082	−0.08 to 0.24	.32

Table . Statistical analysis of weekly questionnaires.

	Participant						
	1	2	3	4	5	6	7
PSEQ[a]							
Pretreatment, mean (SD)	43 (0.7)	31 (3.5)	42 (4.2)	21 (8.5)	37 (4.9)	23 (2.8)	27 (0)
Posttreatment, mean (SD)	38 (2.8)	36 (0)	47 (2.1)	23 (2.1)	45 (1.4)	18 (2.1)	29 (3.5)
RCI[b]	−1.05	1.05	1.05	0.42	1.68	−1.05	0.42
CPAQ[c]							
Pretreatment, mean (SD)	23 (0)	32 (0.7)	31 (0.7)	20 (0.7)	29 (1.4)	15 (1.4)	18 (5.7)
Posttreatment, mean (SD)	28 (1.4)	31 (3.5)	31 (2.8)	23 (0)	29 (1.4)	15 (2.1)	20 (2.1)
RCI	0.74	−0.15	0	0.45	0	0	0.30
PCI[d] active							
Pretreatment, mean (SD)	31 (0.7)	31 (1.4)	31 (1.4)	29 (0.7)	26 (0.7)	28 (0.7)	30 (1.4)
Posttreatment, mean (SD)	28 (1.4)	28 (0)	34 (0)	26 (0)	27 (2.8)	23 (1.4)	30 (0.7)
RCI	−0.84	−0.84	0.84	−0.84	0.28	−1.40	0
PCI passive							
Pretreatment, mean (SD)	40 (1.4)	44 (5.7)	42 (0)	64 (2.8)	46 (3.5)	49 (0.7)	51 (4.2)
Posttreatment, mean (SD)	43 (4.2)	44 (0.7)	36 (.7)	59 (1.4)	44 (0.7)	45 (0)	55 (1.4)
RCI	−0.38	0	0.77	0.64	0.26	0.51	−0.51

[a]PSEQ: Pain Self-Efficacy Questionnaire.

[b]RCI: Reliable Change Index.

[c]CPAQ: Chronic Pain Acceptance Questionnaire.

[d]PCI: Pain Coping Inventory.

Focus Group Analysis

Participants 4, 6, and 7, as described in Table 2, participated in the postintervention focus group. The other participants were not able to participate because they were too busy (with their pain rehabilitation program) (n=3), and did not feel well on the day of the focus group (n=1). Based on the analysis of the focus group, the following three themes were identified: (1) experiences of CMP patients with VR, (2) feasibility of VR, and (3) VR in CMP rehabilitation.

Theme 1: Experiences of CMP Patients With VR

Participants found the VR program attractive to use and valued the intuitive nature of the intervention. Furthermore, they reported several positive effects of the VR intervention, including feelings of self-efficacy, more knowledge about (chronic) pain and focus shifting. Although, these effects were not substantial and patients had to get used to using VR, as it demanded both their time and effort.

> *And it provided me with insights about how chronic pain works.* [Participant 7]
>
> *My focus shifted away from the pain and went more towards the game or killing those monsters, which was a lot of fun. And then you notice that it does something with the pain.* [Participant 6]
>
> *And then you still [use VR] while you are actually already tired and in need of a bit of a rest.* [Participant 4]

Theme 2: Feasibility of VR

Participants perceived the VR intervention as feasible. They found it easy and comfortable to use at home, the instructions were clear, and it was attainable to use daily.

> *And we received clear instructions beforehand, so then it's just plug and play, you know.* [Participant 4]
>
> *Yes, I think I actually liked using it at home first, instead of somewhere else.* [Participant 6]

Theme 3: VR in CMP Rehabilitation

VR helped participants bridge the waiting time, but participants valued it more as an addition to their treatment rather than a substitution.

> *It's more of an addition, a good addition, a meaningful addition.* [Participant 6]

Some participants mentioned it might be valuable to provide the VR intervention not only during the waiting list period but also during the pain treatment they were on the waiting list for. Furthermore, it is important to consider the individual process and whether a patient is open to working on the topics addressed in the VR intervention.

> *…that it would be even more effective during pain treatment, it would be even stronger, because you are already more involved in it and you can also ask for feedback immediately, for example from one of your therapists, if you have any questions.* [Participant 7]

> *It [the VR intervention] raised some internal conflict, but I can really understand that it could be very helpful for patients who are further in their process.* [Participant 4]

In the future, patients would recommend to receive VR not on a daily basis, but maybe 2 or 3 times a week, in between the days of the pain rehabilitation program.

Discussion

Principal Findings

The aim of this study was to gain insight into the influence of VR on pain-related variables and evaluate the feasibility and general experience of this intervention. Analyses of the reported measures showed no clinical and statistically significant differences. Our results imply that the provided intervention did not influence the outcome measures used in this study. This was supported by the visual analyses, which showed that some participants somewhat improved after the intervention on several outcome measures, but worsened on different outcome measures. However, results of the focus group showed that patients qualitatively reported a positive perspective and experienced the intervention as feasible.

Comparison to Previous Work

The results of this study are comparable to other studies that provided the VR intervention, Reducept. A previous study that examined the effect of Reducept for patients with CLBP who were on a waiting list to receive pain treatment [9], showed no significant between-group results on the primary and most other outcome measures, except for opioid use, daily worst, and least experienced pain intensity. It should be noted that the patient sample in both their and our study were patients with severe and complex symptoms. They were referred to secondary pain care, with for example a median pain duration of 5 years in our sample. Previous studies showed that a longer duration of pain complaints was associated with a worse prognosis [43,44] and diminished responsivity to treatment [45]. As suggested before, this specific stand-alone VR intervention might therefore be more suitable for CMP patients with less complex complaints [17].

This study by de Vries et al [17] found somewhat more promising results when they conducted a SCED study among patients with CLBP where they received 9 to 12 45-minute sessions of the VR intervention [17]. Results of their study showed that Reducept might be able to induce clinically relevant reductions in pain intensity and other pain-related outcomes in some patients [17]. These patients were not on a waiting list to receive other pain treatment and received the intervention supervised in the hospital, which might have increased effectiveness [46]. Other interventions that used a stand-alone at-home VR intervention reported clinically meaningful results [47-49], but patients were (1) not on a waiting list to receive other pain treatment and (2) received a more extensive intervention (both in duration and content). A waiting list period is known to possibly deteriorate pain complaints [8]. A meta-analysis among psychotherapies even showed that waiting lists might be regarded as a nocebo condition since patients might, for example, feel the need to remain their complaints to be able to start the pain treatment they are on the waiting list for [50]. In addition, it might be possible that the waiting list period is not the best time to provide VR. This was mentioned in our focus group, and previous research showed that it is also possible to extend secondary care for CMP patients with VR as an additional treatment option [51,52]. In regard to the content of the VR module, it might be possibile to supplement this with, for example, personalized exercise therapy as was done in previous VR interventions for CMP [51,53,54]. Finally, the dosage of the VR intervention might be a point of interest, as the study by de Vries et al [17] found different results from this study while using another dosage of the same intervention. The intervention duration in this trial was 4 weeks, while for behavioral CMP interventions, a duration of 6 to 10 weeks is advised [55], which implies that the intervention did not last long enough. Future studies on VR for CMP should, therefore, study the optimal timing, (personalized) content, and dosage of VR interventions for the most fitting patients.

Results of our study showed a discrepancy between the analyses of quantitative outcome measures and qualitative measures. This is congruent with the qualitative evaluation [22] of the trial that was discussed before [9]. They reported that the VR intervention positively affected how patients' health was experienced, provided patients with more control over their pain, and helped patients accept and understand pain. This is supported by other studies in which patients did not report significant differences in, for example, quality of life or pain intensity measured using questionnaires but mentioned positive benefits during an oral evaluation after their VR intervention [17,56]. This discrepancy could partially be explained by social-desirability bias, as patients might want to portray a more positive impression of the intervention for the researcher who is interviewing them [57]. In addition, it might be possible that nonoptimal quantitative outcome measures were used for this VR intervention, and softer outcomes like values (eg, autonomy) or more proximate outcomes (eg, knowledge about CMP) should be examined as well, as was suggested previously [14].

Strengths and Limitations

One of the strengths of this study was the use of a heterogeneous sample of patients with ranging ages (31-61 years), pain duration (1-30 years), and type of pain complaints. In addition, a rich dataset with multiple subjective (ie, daily diary, validated questionnaires, and focus group) and objective (ie, wearable) outcome measures was used, which was analyzed both visually and statistically. In line with SCED study recommendations, at least 5 data points per phase were collected [58].

This study had several limitations. First, the nature of the study design is characterized by a smaller sample size, which came with risks of selection-bias of specific patients and hindered generalizability of study results. Second, treatment fidelity varied between participants, and not all participants used the VR intervention as much as prescribed, which could have diminished the intervention effect. This problem was mentioned in other VR interventions for CMP as well [48,53], while it is known that repetition is key in, for example, PNE [59]. However, it should be noted that treatment fidelity varies outside a study design, and therefore, this study reflects a real-world situation. Third, we conducted only 1 focus group with 3 participants who provided an insight into the intervention feasibility. Given the limited sample size, these results should be interpreted with caution. However, a more in-depth analysis of qualitative data, possibly with one-on-one interviews instead of focus groups, of participants' experience with VR in a larger study sample would be interesting, to learn more about possible working mechanisms and administration best practices of VR for CMP, which could further improve this intervention.

Future Directions

The results of this study suggest implications for clinical and theoretical practice. It seems that this stand-alone VR intervention for patients with CMP on a waiting list for secondary care does not influence pain-related complaints. However, in the right dose, setting, and timing it might be more effective, as previous research, for example, suggested that VR interventions for CMP might be more effective for younger patients [60]. To further inform trial and intervention design, other relevant pain-related outcomes (eg, catastrophizing) and medication use could be investigated, as these were found relevant in previous VR for CMP studies [9]. In addition, future studies could explore prognostic patient characteristics to identify patients who would respond better or worse to therapeutic VR for CMP. To further study the effectiveness of the (improved) intervention and complement the findings of this study, a randomized controlled trial (RCT) is warranted, in which a control group that receives usual care should be included. This RCT should both focus on the short-term results and include an analysis of the complete pain treatment trajectory. Furthermore, subgroup analyses are needed to examine for which patients VR is effective.

The results of this study showed that this stand-alone immersive VR intervention for patients with CMP on a waiting list did not seem to alter pain-related outcomes. Patients reported good feasibility and general positive experience of the intervention and these outcomes can inform further intervention and trial design.

Acknowledgments

The authors would like to thank all patients who participated in this study and Martijn Eenhoorn and Job Brinkman for their help in data collection. This study was funded by the 2021 Pioneers in Healthcare innovation fund. The funder had no role in the design, organization, and execution of the study.

Data Availability

The datasets generated during this study will not be publicly available but will be available upon reasonable request to the corresponding author.

Authors' Contributions

SS was the principal investigator of this study and drafted the first version of the manuscript. LH conceptualized and designed the study, reviewed and revised the manuscript, and performed supervision. SS, RA, JB, NMDO, RTR, and MS supported recruitment of patients and reviewed and revised the manuscript. MT conceptualized and designed the study, reviewed and revised the manuscript, and supervised SS. All authors contributed to the manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Single-Case Reporting Guideline in Behavioural Interventions (SCRIBE) checklist.
[DOCX File, 19 KB - xr_v2i1e58784_app1.docx]

Multimedia Appendix 2
Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist.
[DOCX File, 26 KB - xr_v2i1e58784_app2.docx]

Multimedia Appendix 3
Topic list focus group.
[DOCX File, 13 KB - xr_v2i1e58784_app3.docx]

Multimedia Appendix 4
Visual analysis of diary data on physical and emotional functioning.
[DOCX File, 242 KB - xr_v2i1e58784_app4.docx]

Multimedia Appendix 5
Clearer version of “Visual analysis of diary data on pain intensity.”
[PPTX File, 118 KB - xr_v2i1e58784_app5.pptx]

Multimedia Appendix 6
Clearer version of “Visual analysis of diary data on pain interference.”
[PPTX File, 124 KB - xr_v2i1e58784_app6.pptx]

Multimedia Appendix 7
Individual scores on steps, stress, and sleep.
[DOCX File, 371 KB - xr_v2i1e58784_app7.docx]

Multimedia Appendix 8
Group scores on weekly questionnaires.
[DOCX File, 17 KB - xr_v2i1e58784_app8.docx]

References

1. Bekkering GE, Bala MM, Reid K, et al. Epidemiology of chronic pain and its treatment in The Netherlands. Neth J Med 2011 Mar;69(3):141-153. [Medline: 21444943]
2. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
3. Hadi MA, McHugh GA, Closs SJ. Impact of chronic pain on patients' quality of life: a comparative mixed-methods study. J Patient Exp 2019 Jun;6(2):133-141. [doi: 10.1177/2374373518786013]
4. Hooten WM. Chronic pain and mental health disorders: shared neural mechanisms, epidemiology, and treatment. Mayo Clin Proc 2016 Jul;91(7):955-970. [doi: 10.1016/j.mayocp.2016.04.029] [Medline: 27344405]
5. Hylands-White N, Duarte RV, Raphael JH. An overview of treatment approaches for chronic pain management. Rheumatol Int 2017 Jan;37(1):29-42. [doi: 10.1007/s00296-016-3481-8] [Medline: 27107994]
6. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021 May 29;397(10289):2082-2097. [doi: 10.1016/S0140-6736(21)00393-7] [Medline: 34062143]
7. Melzack R. From the gate to the neuromatrix. PAIN 1999 Aug;Suppl 6:S121-S126. [doi: 10.1016/S0304-3959(99)00145-1] [Medline: 10491980]
8. Lynch ME, Campbell F, Clark AJ, et al. A systematic review of the effect of waiting for treatment for chronic pain. Pain 2008 May;136(1-2):97-116. [doi: 10.1016/j.pain.2007.06.018] [Medline: 17707589]
9. Groenveld TD, Smits MLM, Knoop J, et al. Effect of a behavioral therapy-based virtual reality application on quality of life in chronic low back pain. Clin J Pain 2023 Jun 1;39(6):278-285. [doi: 10.1097/AJP.0000000000001110] [Medline: 37002877]
10. McCloy R, Stone R. Science, medicine, and the future. Virtual reality in surgery. BMJ 2001 Oct 20;323(7318):912-915. [doi: 10.1136/bmj.323.7318.912] [Medline: 11668138]
11. Trost Z, France C, Anam M, Shum C. Virtual reality approaches to pain: toward a state of the science. PAIN 2021;162(2):325-331. [doi: 10.1097/j.pain.0000000000002060]
12. Liu Z, Wangluo S, Dong H. Advances and tendencies: A review of recent studies on virtual reality for pain management. In: Lackey S, Shumaker R, editors. Virtual Augment Mix: Springer International Publishing; 2016:512-520. [doi: 10.1007/978-3-319-39907-2_49]
13. Hamaker EL. The curious case of the cross-sectional correlation. Multivariate Behav Res 2024;59(6):1111-1122. [doi: 10.1080/00273171.2022.2155930] [Medline: 36599045]
14. Smits M, Ludden GDS, Verbeek PP, van Goor H. How digital therapeutics are urging the need for a paradigm shift: from evidence-based health care to evidence-based well-being. Interact J Med Res 2022 Oct 20;11(2):e39323. [doi: 10.2196/39323] [Medline: 36264624]
15. Morley S. Single Case Methods in Clinical Psychology: A Practical Guide: Routledge; 2017.

16. Smith JD. Single-case experimental designs: a systematic review of published research and current standards. Psychol Methods 2012 Dec;17(4):510-550. [doi: 10.1037/a0029312] [Medline: 22845874]
17. de Vries FS, van Dongen RTM, Bertens D. Pain education and pain management skills in virtual reality in the treatment of chronic low back pain: A multiple baseline single-case experimental design. Behav Res Ther 2023 Mar;162:104257. [doi: 10.1016/j.brat.2023.104257] [Medline: 36731183]
18. Kratochwill TR, Hitchcock JH, Horner RH, et al. Single-case intervention research design standards. Remedial Spec Educ 2013 Jan;34(1):26-38. [doi: 10.1177/0741932512452794]
19. Tate RL, Perdices M, Rosenkoetter U, et al. The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 statement. Phys Ther 2016 Jul;96(7):e1-e10. [doi: 10.2522/ptj.2016.96.7.e1] [Medline: 27371692]
20. Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457. [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
21. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042]
22. Smits M, van Goor H, Kallewaard JW, Verbeek PP, Ludden GDS. Evaluating value mediation in patients with chronic low-back pain using virtual reality: contributions for empirical research in Value Sensitive Design. Health Technol (Berl) 2022;12(4):765-778. [doi: 10.1007/s12553-022-00671-w] [Medline: 35505793]
23. Slatman S, Ostelo R, van Goor H, Staal JB, Knoop J. Physiotherapy with integrated virtual reality for patients with complex chronic low back pain: protocol for a pragmatic cluster randomized controlled trial (VARIETY study). BMC Musculoskelet Disord 2023 Feb 20;24(1):132. [doi: 10.1186/s12891-023-06232-0] [Medline: 36803315]
24. Turk DC, Dworkin RH, Allen RR, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003;106(3):337-345. [doi: 10.1016/j.pain.2003.08.001]
25. Overton M, Ward S, Swain N, et al. Are ecological momentary assessments of pain valid and reliable? a systematic review and meta-analysis. Clin J Pain 2023 Jan;39(1):29-40. [doi: 10.1097/AJP.0000000000001084]
26. Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. Eur J Pain 2007 Feb;11(2):153-163. [doi: 10.1016/j.ejpain.2005.12.008] [Medline: 16446108]
27. Fish RA, McGuire B, Hogan M, Morrison TG, Stewart I. Validation of the chronic pain acceptance questionnaire (CPAQ) in an Internet sample and development and preliminary validation of the CPAQ-8. Pain 2010 Jun;149(3):435-443. [doi: 10.1016/j.pain.2009.12.016] [Medline: 20188472]
28. Kraaimaat FW, Bakker ABH, Evers AWM. Pijncoping-strategieën bij chronische pijnpatiënten: de ontwikkeling van de pijn-coping-inventarisatielijst (PCI) [article in dutch]. Radboud University. 1997. URL: https://repository.ubn.ru.nl/handle/2066/24661 [accessed 2023-10-05]
29. Kraaimaat FW, Evers AWM. Pain-coping strategies in chronic pain patients: psychometric characteristics of the pain-coping inventory (PCI). Int J Behav Med 2003;10(4):343-363. [doi: 10.1207/s15327558ijbm1004_5] [Medline: 14734263]
30. Maas L, Vet H, Köke A, Bosscher RJ, Peters ML. Psychometric properties of the Pain Self-Efficacy Questionnaire (PSEQ): Validation, prediction, and discrimination quality of the Dutch version. Eur J Psychol Assess 2012;28(1):68-75. [doi: 10.1027/1015-5759/a000092]
31. Reneman MF, Dijkstra A, Geertzen JHB, Dijkstra PU. Psychometric properties of Chronic Pain Acceptance Questionnaires: A systematic review. Eur J Pain 2010 May;14(5):457-465. [doi: 10.1016/j.ejpain.2009.08.003]
32. Forerunner® 255 series owner's manual. Garmin. 2022. URL: https://www8.garmin.com/manuals/webhelp/GUID-676967A0-1B23-4384-9BC9-76F3D643F1C8/EN-US/Forerunner_255_OM_EN-US.pdf [accessed 2023-01-26]
33. Verhage F. Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar [report in Dutch]. : Koninklijke Van Gorcum; 1964.
34. Manolov R, Gast DL, Perdices M, Evans JJ. Single-case experimental designs: Reflections on conduct and analysis. Neuropsychol Rehabil 2014 Jul 4;24(3-4):634-660. [doi: 10.1080/09602011.2014.903199]
35. Tate R, Perdices M. Single-Case Experimental Designs for Clinical Research and Neurorehabilitation Settings: Planning, Conduct, Analysis and Reporting: Routledge; 2018. [doi: 10.4324/9780429488184]
36. Bulté I, Onghena P. When the truth hits you between the eyes. Methodology (Gott) 2012 Aug;8(3):104-114. [doi: 10.1027/1614-2241/a000042]
37. Bulté I, Onghena P. The single-case data analysis package: analysing single-case experiments with R software. J Mod App Stat Meth 2013;12(2):450-478. [doi: 10.22237/jmasm/1383280020]
38. Parker RI, Vannest KJ, Davis JL. Effect size in single-case research: a review of nine nonoverlap techniques. Behav Modif 2011 Jul;35(4):303-322. [doi: 10.1177/0145445511399147]
39. Vannest KJ, Parker RI, Gonen O, Adiguzel T. Single Case Research: web based calculators for SCR analysis. Single Case Research. 2011. URL: https://singlecaseresearch.org/calculators/ [accessed 2025-01-24]
40. Jacobson NS, Truax P. Methodological issues & strategies in clinical research. In: Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research: American Psychological Association; 1992:631-648. [doi: 10.1037/10109-042]
41. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94(2):149-158. [doi: 10.1016/S0304-3959(01)00349-9]

42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. Henschke N, Maher CG, Refshauge KM, et al. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. BMJ 2008 Jul 7;337:a171. [doi: 10.1136/bmj.a171] [Medline: 18614473]
44. Von Korff M, Miglioretti DL. A prognostic approach to defining chronic pain. Pain 2005;117(3):304-313. [doi: 10.1016/j.pain.2005.06.017]
45. Wang CK, Myunghae Hah J, Carroll I. Factors contributing to pain chronicity. Curr Pain Headache Rep 2009 Feb;13(1):7-11. [doi: 10.1007/s11916-009-0003-3] [Medline: 19126364]
46. Matthias MS, Bair MJ. The patient–provider relationship in chronic pain management: where do we go from here? Pain Med 2010 Dec;11(12):1747-1749. [doi: 10.1111/j.1526-4637.2010.00998.x]
47. Garcia LM, Birckhead BJ, Krishnamurthy P, et al. An 8-Week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. J Med Internet Res 2021 Feb 22;23(2):e26292. [doi: 10.2196/26292] [Medline: 33484240]
48. Maddox T, Oldstone L, Sparks CY, et al. In-home virtual reality program for chronic lower back pain: a randomized sham-controlled effectiveness trial in a clinically severe and diverse sample. Mayo Clinic Proceedings: Digital Health 2023 Dec;1(4):563-573. [doi: 10.1016/j.mcpdig.2023.09.003]
49. Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-administered skills-based virtual reality intervention for chronic pain: randomized controlled pilot study. JMIR Form Res 2020 Jul 7;4(7):e17293. [doi: 10.2196/17293] [Medline: 32374272]
50. Furukawa TA, Noma H, Caldwell DM, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta‐analysis. Acta Psychiatr Scand 2014 Sep;130(3):181-192. [doi: 10.1111/acps.12275]
51. Ummels D, Cnockaert E, Timmers I, den Hollander M, Smeets R. Use of virtual reality in interdisciplinary multimodal pain treatment with insights from health care professionals and patients: action research study. JMIR Rehabil Assist Technol 2023 Nov 10;10:e47541. [doi: 10.2196/47541] [Medline: 37948109]
52. Goudman L, Jansen J, Billot M, et al. Virtual reality applications in chronic pain management: systematic review and meta-analysis. JMIR Serious Games 2022 May 10;10(2):e34402. [doi: 10.2196/34402] [Medline: 35536641]
53. Tuck N, Pollard C, Good C, et al. Active virtual reality for chronic primary pain: Mixed methods randomized pilot study. JMIR Form Res 2022 Jul 13;6(7):e38366. [doi: 10.2196/38366] [Medline: 35830224]
54. Griffin A, Wilson L, Feinstein AB, et al. Virtual reality in pain rehabilitation for youth with chronic pain: Pilot feasibility study. JMIR Rehabil Assist Technol 2020;7(2):e22620. [doi: 10.2196/22620]
55. Knoerl R, Lavoie Smith EM, Weisberg J. Chronic pain and cognitive behavioral therapy: An integrative review. West J Nurs Res 2016 May;38(5):596-628. [doi: 10.1177/0193945915615869] [Medline: 26604219]
56. Garrett B, Taverner T, McDade P. Virtual reality as an adjunct home therapy in chronic pain management: An exploratory study. JMIR Med Inform 2017 May 11;5(2):e11. [doi: 10.2196/medinform.7271] [Medline: 28495661]
57. Bergen N, Labonté R. “Everything Is Perfect, and We Have No Problems”: Detecting and limiting social desirability bias in qualitative research. Qual Health Res 2020 Apr;30(5):783-792. [doi: 10.1177/1049732319889354]
58. Ganz JB, Ayres KM. Methodological standards in single-case experimental design: Raising the bar. Res Dev Disabil 2018 Aug;79:3-9. [doi: 10.1016/j.ridd.2018.03.003] [Medline: 29655508]
59. Watson JA, Ryan CG, Cooper L, et al. Pain neuroscience education for adults with chronic musculoskeletal pain: a mixed-methods systematic review and meta-analysis. J Pain 2019 Oct;20(10):1140. [doi: 10.1016/j.jpain.2019.02.011] [Medline: 30831273]
60. Lier EJ, de Vries M, Steggink EM, Ten Broek RPG, van Goor H. Effect modifiers of virtual reality in pain management: a systematic review and meta-regression analysis. PAIN 2023 Aug 1;164(8):1658-1665. [doi: 10.1097/j.pain.0000000000002883] [Medline: 36943251]

Abbreviations

ACT: acceptance and commitment therapy
CBT: cognitive behavioral therapy
CE: Conformité Européenne
CLBP: chronic low back pain
CMP: chronic musculoskeletal pain
COREQ: Consolidated Criteria for Reporting Qualitative Research
COREQ: Consolidated Criteria for Reporting Qualitative Research
CPAQ: Chronic Pain Acceptance Questionnaire
HMD: head-mounted display
HRV: heart rate variability
IMMPACT: Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
PCI: Pain Coping Inventory
PNE: pain neuroscience education

PSEQ: Pain Self-Efficacy Questionnaire
RCI: Reliable Change Index
RCT: randomized controlled trial
SCED: single-case experimental design
SCRIBE: Single-Case Reporting Guideline in Behavioural Interventions
VR: virtual reality

Edited by T Leung; submitted 25.03.24; peer-reviewed by B Fordham; revised version received 02.12.24; accepted 06.12.24; published 04.03.25.

Please cite as:
Slatman S, Heesink L, Achterkamp R, Broeks J, Monteiro de Oliveira N, ter Riet R, Stegeman M, Tabak M
At-Home Virtual Reality Intervention for Patients With Chronic Musculoskeletal Pain: Single-Case Experimental Design Study
JMIR XR Spatial Comput 2025;2:e58784
URL: https://xr.jmir.org/2025/1/e58784
doi:10.2196/58784

Feasibility, Subjective Effectiveness, and Acceptance of Short Virtual Reality Relaxation Breaks for Immediate Perceived Stress Reduction in Emergency Physicians: Single-Arm Pre-Post Intervention Study

Tanja Birrenbach*, MD, MME, PD; Seraina Häni*; Sabrina Jegerlehner, MD, PD; Simon Schober; Aristomenis K Exadaktylos, Prof Dr Med; Thomas C Sauter, MME, Prof Dr Med

Department of Emergency Medicine, Inselspital, University Hospital of Bern, University of Bern, Rosenbühlgasse 27, Bern, Switzerland
*these authors contributed equally

Corresponding Author:
Tanja Birrenbach, MD, MME, PD
Department of Emergency Medicine, Inselspital, University Hospital of Bern, University of Bern, Rosenbühlgasse 27, Bern, Switzerland

Abstract

Background: Emergency physicians face significant stress in their daily work, adversely affecting patient care and contributing to physician burnout.

Objective: This pilot study explored the feasibility, immediate effects, and acceptance of virtual reality (VR) relaxation on perceived stress reduction among emergency physicians.

Methods: The study was conducted at the Department of Emergency Medicine, Bern, Switzerland, in February 2023. All junior and senior physicians were eligible, excluding those with epilepsy, claustrophobia, or severe nausea. Voluntary participants underwent a 6- to 8-minute VR meditation program at their workplace. Subjective short-term stress reduction was measured using a numeric rating scale (NRS) ranging from 0 (“not at all stressed”) to 10 (“extremely stressed”). Feasibility, user acceptance, and technical aspects were evaluated using validated and self-constructed questionnaires.

Results: In total, 35 emergency physicians (median [IQR] age, 32 [30-34] years, 60% female) completed 39 VR simulation sessions. Baseline stress levels (median NRS 4, IQR 2 - 6.5) were significantly reduced post-intervention (median NRS 2, IQR 1 - 4; *P*<.001), particularly among participants with high baseline stress levels. Reported side effects (simulator sickness) were minimal; the median score of presence and immersion according to the questionnaire developed by Slater-Usoh-Steed was 4 (IQR 3 - 4) (scale 1 - 7, with 7=full immersion). User satisfaction was high. Implementation challenges mainly included technical issues and time constraints due to high workload.

Conclusions: This pilot study suggests that brief, relaxing VR sessions may help reduce short-term perceived stress levels in emergency physicians with minimal side effects and high user satisfaction. Future studies should address implementation challenges to optimize integration with clinical workflows.

(JMIR XR Spatial Comput 2025;2:e72605) doi:10.2196/72605

KEYWORDS

virtual reality; relaxation; stress; emergency medicine; workplace; burnout

Introduction

Emergency medicine is an inherently high-stress medical specialty due to the urgent and often severe nature of cases, which demand rapid decision-making with potentially life-altering consequences. The additional burden of shift work and disrupted circadian rhythms further exacerbates stress levels among emergency physicians. These factors contribute to a heightened risk of burnout [1-3], posttraumatic stress disorder [4], substance abuse [5], and even suicide [6]. Burnout is a syndrome conceptualized as resulting from chronic workplace stress that has not been successfully managed. It is characterized by feelings of energy depletion or exhaustion; increased mental distance from one’s job, or feelings of negativism or cynicism related to one’s job; and reduced professional efficacy. A recent Swiss investigation confirmed emergency physicians as a medical specialty at great risk for burnout. Over half of the more than 600 respondents met at least 1 criterion for burnout and reported symptoms of mild to severe depression. Alarmingly, 10% of respondents even reported having considered suicide at some point [7]. The implications of burnout extend beyond individual well-being, jeopardizing patient care quality and

safety and contributing to physicians leaving the profession [7,8]. Therefore, prioritizing personal stress management strategies and advancing research into effective stress reduction methods are essential to maintaining both the quality and sustainability of emergency medicine. This aligns with the World Health Organization's call for addressing health care worker well-being to ensure the resilience of health care systems [9].

Various stress management interventions, such as yoga, mindfulness training, deep breathing exercises, and psychoeducational stress management workshops, have demonstrated effectiveness and are increasingly being implemented in workplace settings. However, the integration of these interventions into fast-paced work environments, such as emergency medicine, remains a significant challenge [10,11].

Virtual reality (VR) is a computer-generated simulation allowing the user to fully immerse himself in an interactive, 3-dimensional environment, typically through a specialized VR headset, or head-mounted device. By blocking out the real world and replacing it with a digital space, VR allows users to engage with virtual objects and environments in real time. This immersion fosters a sense of presence, where users psychologically perceive the virtual world as real, enhancing emotional and cognitive engagement. In relaxation-focused VR applications, this heightened presence allows users to fully disconnect from external stressors, creating a safe space for restorative mental states and stress relief [12].

In the medical field, VR has long been used as a virtual therapeutic tool for managing acute and chronic pain and reducing anxiety across various settings, including the emergency department (ED) [13,14]. Additional applications include treatment for mental health conditions such as cognitive impairment, depression, phobias, and posttraumatic stress [15-17]. Research indicates that VR is an effective therapeutic tool for relaxation, modulating individual stress levels, and potential impacts on the immune response [18]. It offers a cost-effective and accessible option for therapeutic intervention [10,19-21]. Unlike traditional mindfulness practices such as meditation or yoga, VR requires little to no prior experience before positive effects can be achieved [22]. Possible explanations include the attention restoration theory, which posits that exposure to natural environments can replenish cognitive resources depleted by stress. VR can simulate calming natural scenes, providing restorative experiences that reduce mental fatigue and stress [23]. The biopsychosocial model suggests that stress is influenced by biological, psychological, and social factors. VR interventions can address these components by offering immersive experiences that promote relaxation, thereby positively affecting physiological and psychological states.

Potential barriers to the widespread adoption of VR include initial implementation and ongoing maintenance costs, limited accessibility related to hardware availability or user familiarity, uncertainty regarding the duration of beneficial effects, and the risk of adverse reactions such as visually induced motion sickness. Emerging evidence on the use of VR for health care workers suggests promising outcomes. A recent randomized controlled trial involving 32 health care workers demonstrated that VR-based guided meditations are a feasible and accessible mindfulness intervention, potentially even more effective than non-immersive methods [24]. Similarly, brief, tranquil VR experiences have been shown to significantly reduce subjective stress among frontline health care workers during the COVID-19 pandemic [25,26] and to enhance happiness and relaxation among trauma care clinicians [27].

The evidence regarding the use and effectiveness of VR as a stress reduction tool for emergency physicians remains limited. Additionally, implementing VR within the unpredictable and fast-paced environment of an ED presents significant challenges. The feasibility of its application and the acceptance by the emergency team are unclear. Therefore, we conducted a within-subject, repeated measure interventional feasibility pilot study to evaluate the feasibility of deployment of a short relaxing VR simulation in the busy setting of the ED as a stress-reduction tool for emergency physicians; the immediate effect of VR use on self-perceived stress; and the tacceptance of the VR simulation in the study population (user satisfaction, simulator sickness, and sense of presence and immersion).

Methods

Design and Setting

This prospective non-randomized pre-post interventional feasibility pilot study was conducted at the ED of the University Hospital of Bern, Switzerland. As one of the largest EDs in Switzerland, it serves approximately 55,000 patients annually and is staffed by a team of around 70 physicians [28]. The study was carried out between February 1 and February 28, 2023, during daytime hours, contingent on the availability of the study investigators (SH and SS).

The study was conducted on a convenient sample of emergency physicians. Written informed consent was obtained from all participants, including data anonymization and authorization for use in study analysis and publication.

Ethical Considerations

The local ethics committee (Kantonale Ethikkommission Bern, KEK; BASEC number Req-2023 - 00018) classified this study as a quality evaluation project, exempting it from the requirements of the Swiss Human Research Act.

Inclusion and Exclusion Criteria

All junior and senior physicians working in the ED of the University Hospital in Bern were eligible for participation. Exclusion criteria included facial or neck injuries, severe nausea or vomiting, claustrophobia, epilepsy, or any other conditions associated with hypersensitivity to light or motion.

Baseline Data

Baseline data included sociodemographic factors (gender, age), the use of visual aids, and smoking habits. Information regarding work routines was also collected, such as the participant's role in the ED, years of professional experience, board certification, workload percentage, frequency of night shifts per month, average break duration, and typical break activities.

Additionally, participants were asked about prior experience with gaming, VR, and mindfulness exercises (“I regularly use gaming, VR or mindfulness training”). The baseline questionnaire was completed before the initial use of the VR intervention.

Intervention

Physicians were informed about the project in advance during staff meetings, and throughout the study period, reminders were provided through announcements during briefings and informational posters. Additionally, participants were recruited through direct contact by the study coordinators (SH and SS). For half of the 28 days, the study was conducted from 7 AM to 3 PM, and for the other 14 days, from 3 PM to 11 PM, corresponding to the 2 largest daily shifts. During their time at the University Hospital of Bern, the study coordinators were easily reachable via a pager system, allowing physicians to choose an appropriate time for the intervention at their discretion.

The study investigators (SH and SS) informed the participant about the study aims, handed out the information form, ensured the absence of contraindications, responded to the participant’s questions, and collected their free, informed, and expressed consent.

The intervention consisted of the application of a 6- to 8-minute VR relaxation program called “Daily Focus,” including breathing exercises and a short focus exercise in an imaginary environment. The immersive experience consists of a contemplative, relaxing, futuristic imaginary landscape accompanied by a sound universe specifically composed to relax the user. The scenery and theme changed daily. The content also had interactive capabilities as well, so that the user could take action to affect the VR environment. The user could choose to interact with the environment by fixating one’s gaze on an interactive object in the worldscape. “Daily Focus” is part of the commercially available software “TRIPP” developed by TRIPP Inc. (TRIPP Inc.). The company was not involved in any aspects of the study. A commercially available stand-alone head-mounted display (Meta Quest 2; Meta) was used. When it became apparent that background noise at the University Hospital of Bern’s workplaces impaired the sense of immersion for some participants, noise-cancelling headphones (JBL Live 650BTNC; JBL) were introduced to reduce ambient sounds. As the physicians’ experience with VR head-mounted displays was limited, the users were supported by the study team in the technical application when needed (SH and SS). In case of a medical emergency requiring the immediate presence of the physician, the VR simulation was interrupted. The briefing, completion of the consent form and questionnaires, and the intervention itself took approximately 15 minutes in total. The duration of the evaluation and intervention was intentionally kept as short as possible to minimize barriers to participation.

Outcomes

Feasibility

Feasibility was assessed using technical details of the simulation (location of the simulation, ie, directy at the workplace vs quieter location, interruptions of the simulation and reasons for interruptions, and timing of the intervention), as well as with free text comments of the users and feedback collected from the study team (SH and SS).

Immediate Effect of VR Use on Perceived Stress

Perceived stress reduction was measured as the difference between the self-reported stress level directly before and after the intervention on a numeric rating scale (NRS-11) scale from 0 to 10 (0=”not at all stressed” to 10=”extremely stressed”). This simple measure was selected due to its strong correlation with the well-validated Perceived Stress Scale 14 (PSS-14) [29]. Furthermore, a threshold value of 6.8 on the self-reported scale has been shown to effectively predict high stress levels, corresponding to a PSS-14 cutoff score of ≥7.2, and was therefore chosen to identify individuals experiencing high stress, similar to Beverly et al [25,29].

User Acceptance

User acceptance was evaluated using the following questionnaires:

Visually induced motion sickness was assessed according to the Simulator Sickness Questionnaire (SSQ) from Kennedy et al [30].

Presence and immersion in the virtual world were determined according to the 6-item questionnaire developed by Slater-Usoh-Steed (total score ranges from 1=no immersion to 7=full immersion) [31].

User satisfaction was assessed using a self-constructed 8-item questionnaire (1: I enjoyed the simulation experience; 2: The headset and headphones felt comfortable; 3: The audio quality was clear and enjoyable; 4: The image quality was visually pleasing; 5: The simulation helped to reduce my stress level; 6: I would use this simulation again for relaxation; 7: I would recommend this simulation to others; 8: The simulation can be conveniently performed directly at the workplace). Responses were collected on a 5-point Likert scale (1=“totally disagree” to 5=“totally agree”) immediately following the intervention.

Furthermore, a self-constructed 6-item user acceptance questionnaire was sent out 2 weeks after the final intervention via email to all physicians working in the department, with a particular focus on understanding the limiting factors that prevented users from taking a break with VR (1: I couldn’t find time during my shift because the workload was too high; 2: I felt it wasn’t worth investing the time because I preferred to finish my documentation as early as possible to end my shift on time; 3: I didn’t enjoy the simulation (virtual environment/voice guidance), but I could imagine using it more often with a different program; 4: I experienced side effects that overshadowed the positive aspects of the VR breaks; 5: I prefer to spend my breaks differently; 6: I didn’t think about it/forgot that the option was available. Responses were collected on a 5-point Likert scale (1=“totally disagree” to 5=“totally agree”).

Statistical Analysis

Statistical analysis was carried out using Python (version 3.9.12) and the following packages: NumPy, SciPy (matplotlib, seaborn). Baseline characteristics are presented as numbers and

percentage or median and interquartile range (IQR) using descriptive statistics as appropriate. Pre- and post-simulation comparisons (stress level) were performed with the Wilcoxon signed-rank test.

We performed subgroup analyses, including participants with high stress levels defined as NRS-11 ≥6.8 (similar to Beverly et al [25]) with the Wilcoxon rank sum test.

Comparisons between independent groups (eg, male vs female, status of active patient care involvement, prior experience with mindfulness training, gaming experience) were carried out by Wilcoxon rank sum or Kruskal-Wallis test depending on the variable.

A $P<.05$ was considered significant.

Effect sizes with 95% CI for stress levels before and after the simulation were determined by Cohen *d*. Effect size was determined as follows: Cohen d <0.5 small, 0.5 - 0.8 moderate, and >0.8 large.

Results

Baseline Characteristics

Out of 67 physicians (61% female), 35 working in the ED completed the study (response rate 52.2%). The average age of the participants was 32 (IQR, 30 - 34) years, with 60% (n=21) being female. Further demographic characteristics as well as break behavior are reported in Table 1.

Participants were asked to rate their experience with gaming, VR, and mindfulness training ("I regularly use gaming, VR or mindfulness training") on a scale from 1 ("Strongly disagree") to 5 ("Strongly agree"). For gaming, the median score was 1 (IQR 1 - 2), no participants had prior experience with VR, and for mindfulness training, the median score was 2 (IQR 1 - 3).

Table . Baseline characteristics including break routine (N=35).

Item	Value
Gender, n (%)	
Male	14 (40)
Female	21 (60)
Age in years, median (IQR)	32 (30-34)
Use of visual aids, n (%)	16 (45.7)
Smoker, n (%)	0 (0)
Professional role, n (%)	
Resident physician	27 (77.1)
Fellow physician	2 (5.7)
Senior physician	6 (17.1)
Board certification, n (%)	12 (34.3)
Work experience, years, median (IQR)	5 (4-7)
Employment level, %, median (IQR)	80 (80-100)
Frequency of night shifts per month, median (IQR)	4 (3-5)
Break routine, n (%)	
No breaks	6 (17.1)
Break at the workplace with constant availability	29 (82.9)
Break at the workplace without constant availability	0 (0)
Average break time in minutes, median (IQR)	15 (10–20)

Feasibility and Technical Details of the Interventions

Out of 35 participants, 4 (11.4%) individuals completed the intervention twice, resulting in a total of 39 interventions. The majority of interventions (n=23, 59%) occurred directly at the workplace, while 41% (n=16) took place in designated rooms away from the workplace. Out of 39 interventions, 6 (15.4%) experienced interruptions. The majority of these (66.7%, n=4; 10.3% of all interventions) were due to technical issues, while the remaining 2 (33.3%, 5.1% of all interventions) were caused by urgent medical duties requiring participant attention.

In terms of shift schedules, 61.5% (n=24) of interventions were conducted during the morning shift (7–3 PM), 35.9% (n=14) during the afternoon shift (3 PM–11 PM), and 2.6% (n=1) at the end of a night shift (7 AM). During most of the interventions (n=27, 69.2%), participants remained actively engaged in patient care, whereas about one-third (n=12, 30.8%) were conducted after shift hand-over, with the participants no longer directly

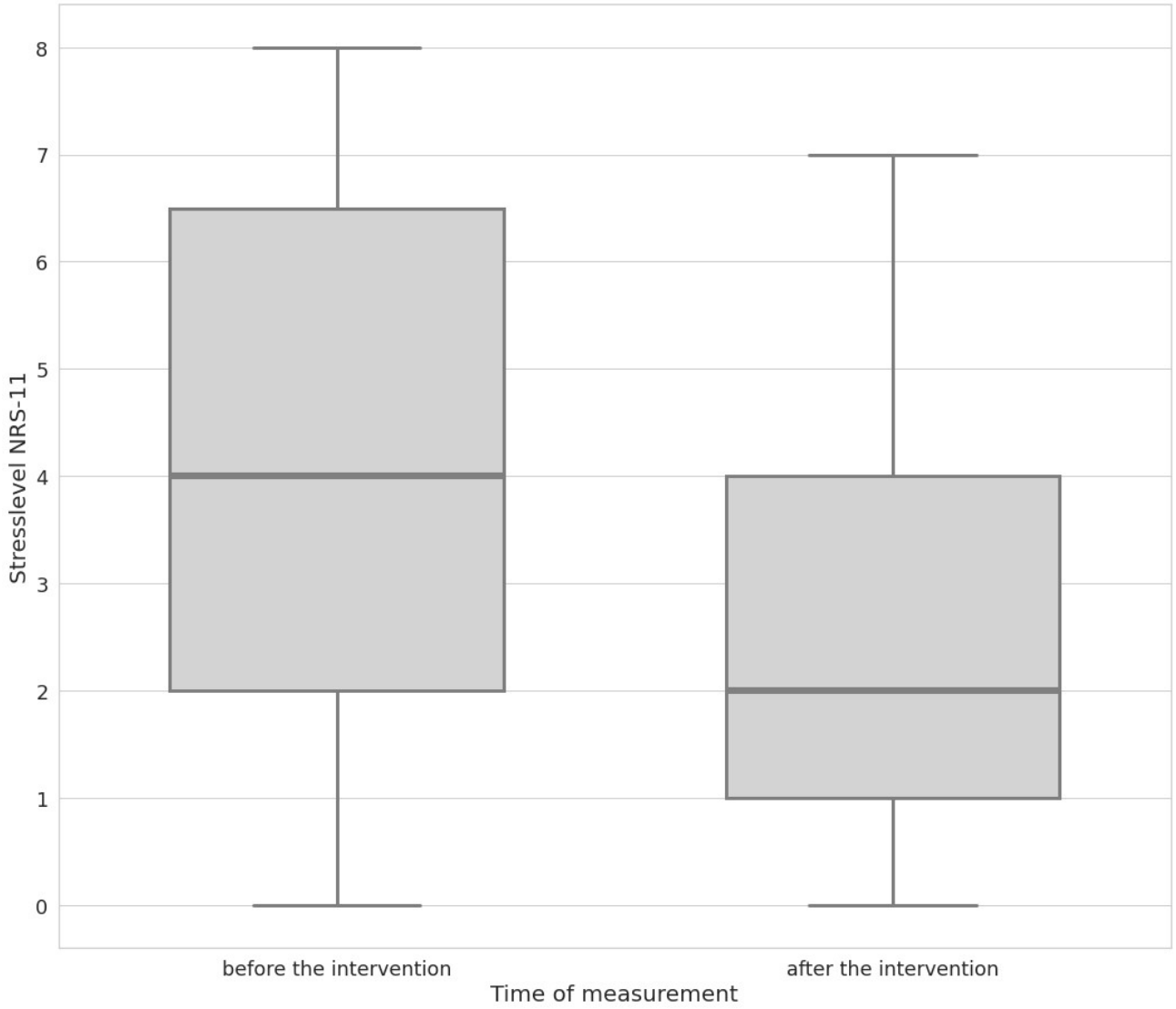
responsible for patient care but remaining engaged in administrative tasks.

The free-text comments were predominantly positive, highlighting the usefulness and effectiveness of the intervention. However, some criticisms were noted regarding the comfort of the headset and aspects of the simulation itself. Suggestions included a more photorealistic scenario and reduced voice guidance during the simulation. Feedback indicating that ambient emergency noises disrupted immersion was addressed by introducing the use of headphones and conducting sessions in quiet, isolated rooms whenever possible. Additionally, participants frequently mentioned that during active patient care, they were often unable to fully engage with the simulation or felt unable to allocate sufficient time for the intervention.

Immediate Effect of VR Use on Perceived Stress Reduction

The baseline median stress level was 4/10 (IQR 2 - 6.5), which was reduced to 2/10 (IQR 1 - 4) after the intervention (P<.001) (Figure 1). The effect size was calculated as Cohen d=1.28 (95% CI 0.84 - 1.72), representing a large effect.

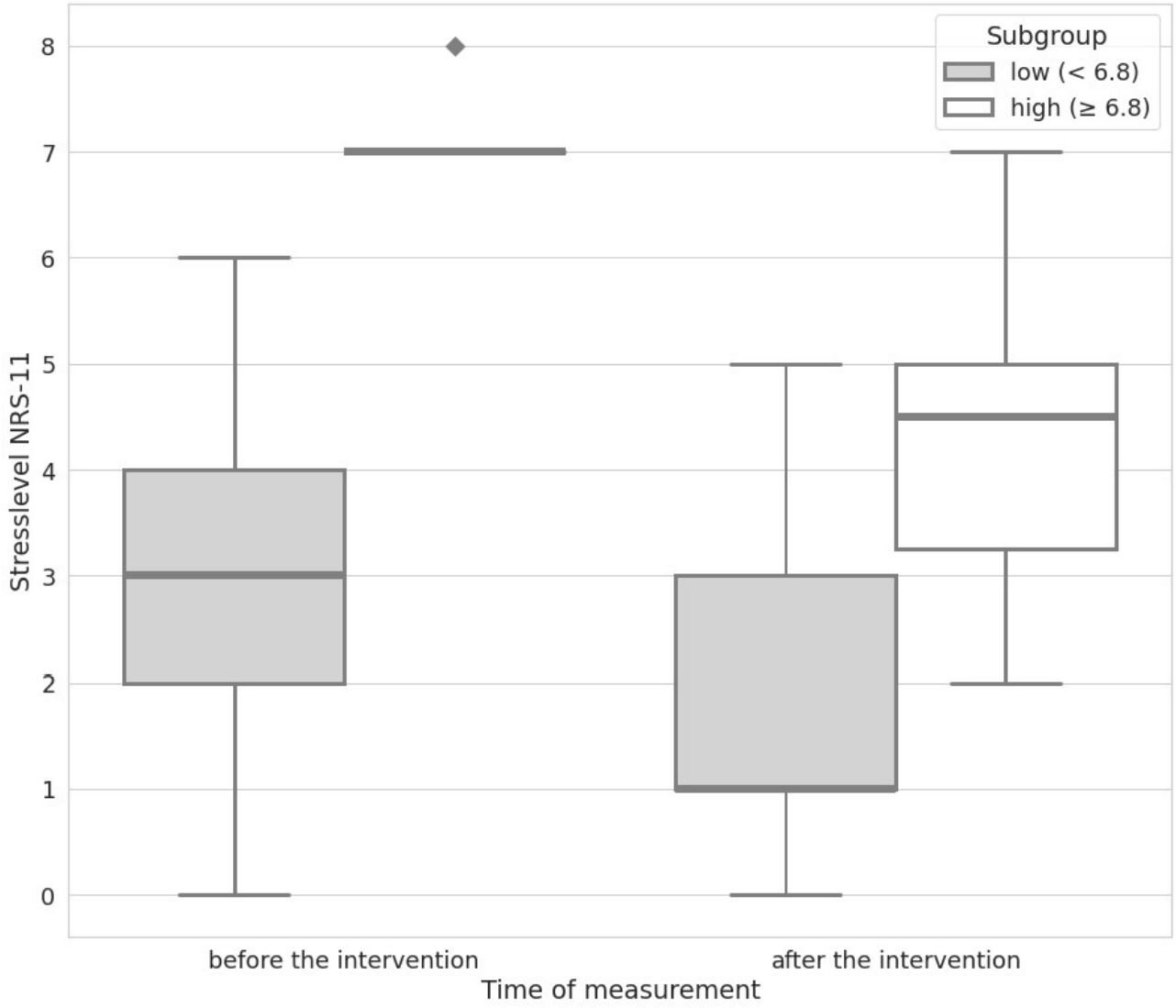
Figure 1. Immediate effect of virtual reality use on perceived stress reduction. Comparison of stress levels before and after the intervention. NRS: numeric rating scale.



In total, 10 participants reported high baseline stress levels (≥6.8). In this group, the intervention was even more effective, reducing the stress level from 7/10 to 4.5/10 (P<.001) (Figure 2). Only one individual reported a high stress level after the intervention. In this case, the simulation was terminated after 2 minutes due to an audio malfunction.

No significant differences in stress reduction concerning the variables gender (P=.767), prior experience with mindfulness training (P=.376), gaming experience (P=.489), or involvement in active patient care (P=.912) were found.

Figure 2. Immediate effect of virtual reality use on perceived stress reduction according to stress level. Comparison of stress levels of subgroups with low and high stress before and after the intervention. Outliers (values ≥1.5×IQR) are indicated as diamonds. NRS: numeric rating scale.



User Acceptance of the VR Simulation

Visually Induced Motion Sickness

The median of the total score according to the SSQ from Kennedy was 80 (IQR 0 - 161) (range 0 - 813).

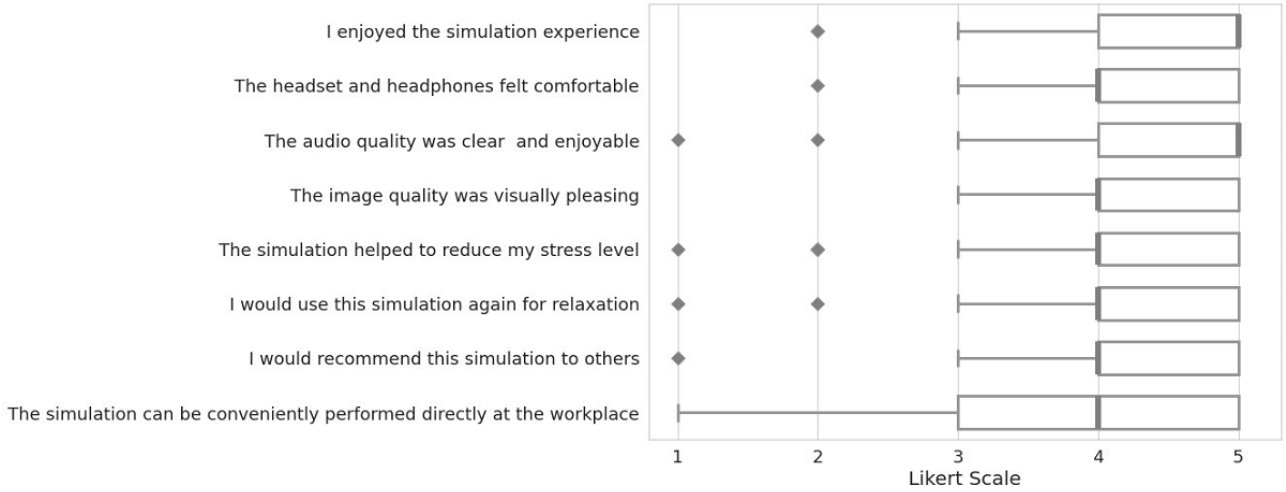
Presence and Immersion

The median score of presence and immersion according to the questionnaire developed by Slater-Usoh-Steed was 4 (IQR 3 - 4) (with 7=full immersion).

User Satisfaction

Results of the user satisfaction survey are detailed in Figure 3.

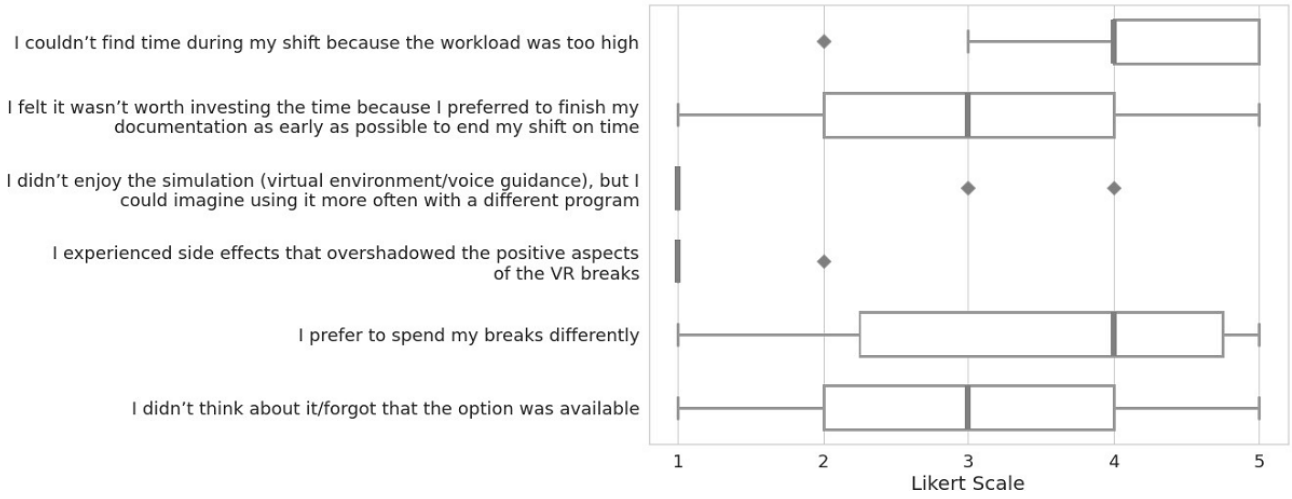
Figure 3. User satisfaction survey. Results of the user satisfaction survey. Answers on a 5-point Likert scale from 1="totally disagree" to 5="totally agree" directly after the intervention. Outliers (values ≥ 1.5×IQR) are indicated as diamonds.



Acceptance Survey

In total, 16 physicians completed the 6-item retrospective acceptance survey sent out 2 weeks after the intervention period (response rate 24%). Answers are depicted in Figure 4.

Figure 4. Retrospective acceptance survey. Results of the retrospective acceptance survey. Answers on a 5-point Likert scale from 1="totally disagree" to 5="totally agree." Outliers (values≥1.5×IQR) are indicated as diamonds.



Discussion

Overview

This pilot study evaluated the feasibility, immediate effect of VR use on perceived stress reduction, and acceptance of VR simulation as a short break intervention within the high-pressure environment of an ED.

We observed a significant reduction in self-reported stress levels, decreasing from 4/10 to 2/10, with a large effect size. Importantly, 26% of participants reported high stress levels prior to the intervention, in which the stress-reducing effect of VR was particularly pronounced. No significant differences in stress reduction were observed across demographic or experiential variables, including gender, prior mindfulness training, gaming experience, or engagement in active patient care during the intervention.

Acceptance was high with minimal side effects. Despite its effectiveness, challenges were noted in implementing VR breaks, primarily due to the substantial time constraints faced by health care professionals. These logistical barriers may limit the practical application of this intervention in routine clinical practice.

Feasibility

The simulation was carried out with minimal technical issues or interruptions. Notably, a significant number of individuals voluntarily participated in the study, and the overall feedback was highly positive, indicating strong interest and acceptance of the concept.

One limitation observed was the execution of the simulation directly in the busy ED environment at the participants' desktops, which occasionally resulted in distracting background noise that could impact participants' focus. To address this issue, noise-canceling headphones were introduced, and approximately

40% of the simulations were conducted in quieter settings in the ED.

Stress Reduction

This study demonstrated a significant reduction in subjective stress levels, with an average decrease of 2 points on the NRS-11 scale following a 6- to 8-minute VR simulation. These findings align with previous studies investigating VR-based stress reduction in high-stress medical environments, particularly during the COVID-19 pandemic.

For example, Beverly et al [25] conducted a similar study involving frontline health care workers. They observed comparable reductions in stress (mean change –2.2 on a visual analogue scale from 1 to 10, effect size Cohen d = 1.08) and high levels of acceptance after a 3-minute 360-degree cine-VR simulation featuring a nature scene. Similarly, Putrino et al [32] reported on the effectiveness of "Recharge Rooms," immersive multisensory environments designed to alleviate stress among frontline health care workers during the pandemic. These rooms incorporated visual projections of natural landscapes, calming sounds, and soothing scents. In a study involving 496 participants, average self-reported stress scores decreased significantly from 4.58 to 1.85 on a 6-point scale after a single 15-minute session, with high user satisfaction reported. Further supporting these findings, Nijland et al [26] evaluated the use of 10-minute VR relaxation breaks in 360° immersive environments for 86 ICU nurses during their shifts. This intervention demonstrated similar reductions in stress levels (mean change –1.4 on a visual analogue scale from 1 to 10) and high user acceptance. However, a key barrier identified across studies, consistent with our findings, was the high workload of health care professionals, which limited the feasibility of integrating VR-based interventions into routine clinical practice.

While no studies specifically targeted the ED setting, Adhyaru and Kemp [27] reported on the use of VR relaxation interventions among 39 predominately female physicians working in a fast-paced trauma service. The study highlighted the positive impact of 10-minute VR relaxation sessions using the Nature Treks application, demonstrating the potential for VR-based interventions in similar high-stress environments. Participants engaged in these sessions within a designated well-being room during their workday, immersing themselves in natural environments. Post-intervention, participants reported significant increases in feelings of happiness and relaxation, accompanied by notable decreases in sadness, anger, and anxiety. Objective measures also showed a significant reduction in heart rate, indicating decreased physiological arousal.

Although the short-term effects of various VR applications appear comparable, meaningful comparisons remain challenging due to differences in study settings, target populations, specific content and design of VR software, as well as external factors such as the surrounding environment and circumstances (eg, pandemic conditions). These variations significantly limit the generalizability and interpretability of findings across different VR studies.

Speculatively, VR's effectiveness might be attributed to attention restoration theory, proposing that immersive restorative environments help replenish cognitive resources depleted by stress [23]. Additionally, the biopsychosocial model posits that immersive VR experiences can modulate neurophysiological responses, such as decreasing sympathetic nervous system activation and reducing cortisol levels, thereby alleviating stress [18,21].

Nevertheless, the optimal design of VR-based stress interventions remains unclear. Current literature varies widely regarding realism (naturalistic vs abstract scenarios), activity levels (passive viewing vs interactive tasks), and intervention types (guided meditation vs free exploration). These variations underline the necessity for further research using rigorous experimental designs with both cognitive and neurophysiological methodologies. Future studies should systematically investigate these variables to identify the most effective VR intervention formats and better elucidate the underlying mechanisms driving VR-induced stress reduction.

As this was a pilot study, only short-term (pre-post) effects regarding stress reduction were evaluated. However, findings from several studies provide initial data supporting the effectiveness of long-term VR-based programs for reducing stress, anxiety, and burnout among different health care professionals [24,33-36]. A recent study in the ED explored the effectiveness of a 4-week VR-based mindfulness intervention using brief guided breathing exercises. Participants using VR demonstrated greater improvements in relaxation, as measured by heart rate variability (HRV), compared to a mobile app. Regular VR use led to increased relaxation effectiveness over time, suggesting VR's suitability for long-term mindfulness programs [24]. Several aspects require further study, such as examining patterns in VR mindfulness effectiveness across varying workload conditions and shifts. Additionally, stress and relaxation trends could be assessed by demographic or professional differences like job role or experience. It would also be valuable to explore the cumulative impact of VR sessions on chronic stress and burnout over time, analyze the timing of sessions related to well-being outcomes, and investigate how individual personality traits or baseline stress resilience influence responses to VR interventions.

User Acceptance

With regard to side effects, the intervention proved to be largely free of adverse effects. This aligns with findings from other studies that have used VR as a relaxation tool [14,36,37].

The results indicated only moderate levels of immersion, consistent with findings from another study investigating the use of VR for pain reduction in our ED setting [14]. For both studies, we attribute this moderate immersion to environmental factors such as background noise, interruptions, and the generally high-stress atmosphere. These factors likely relate to the aforementioned limitations in implementing VR interventions within the workplace.

Overall, user satisfaction among participants was very high. Comfort, as well as the audio and visual quality, received considerable praise, particularly after the introduction of noise-canceling headphones. Participants also reported high subjective effectiveness for relaxation, with strong agreement

on statements such as, “I would use this simulation again for relaxation” and “I would recommend this simulation to others.”

However, the statement “The simulation can be easily conducted in the workplace” received less agreement. This raises the question of whether and how the intervention could be better integrated into the ED workplace setting. As revealed by the retrospective questionnaire, many participants did not engage in the intervention due to work-related time pressures. This highlights a broader issue also reflected in the baseline survey results. On average, physicians reported taking only 15 minutes for breaks during their shifts.

The strong agreement with the statement “I couldn’t find time during my shift because the workload was too high” further underscores a structural challenge related to workload and break culture within the workplace. Spontaneous comments from participants and the low-medium baseline stress levels suggest that participants only took time for the intervention after the peak of their stress had passed. Given that the highest levels of work-related stress for emergency physicians typically occur during the care of critically ill patients, this timing is likely unavoidable—and perhaps even desirable. As highlighted in a recent phenomenographic study on well-being interventions in the ED, the demands of the job simultaneously necessitate and limit the implementation of effective interventions to support staff well-being in this challenging environment [38]. Possible solutions for further interventions include protected break and VR break times or scheduling VR breaks during lower workload periods.

Some participants criticized the fantasy-style design of the simulation, expressing a preference for a more naturalistic environment. However, the software used has been successfully applied in several other settings [21,34,37]. Meanwhile, many studies investigating VR for stress reduction have used realistic nature-based simulations, such as a forest. Such an approach may further enhance relaxation, as numerous studies have demonstrated that exposure to forests and nature in general promotes relaxation [10,18,20].

While we demonstrated technical feasibility and user acceptance of short VR interventions, factors such as device affordability, software licensing costs, and the scalability of deploying VR systems across various clinical settings must be carefully considered.

Limitations

This study has several limitations that should be considered when interpreting the results. First, and mainly, the absence of a control group makes it impossible to definitively attribute the observed stress reduction to the VR intervention itself. Without a comparator, we cannot rule out alternative explanations, such as placebo effects, spontaneous recovery, or other external factors. However, given the feasibility nature of this pilot study and the promising results observed, these findings provide a solid foundation for future controlled trials. These should incorporate a more rigorous design, eg, a randomized controlled trial with a control group or an active control condition (eg, a non-VR relaxation technique, like guided breathing exercises). Second, no other structured assessments for burnout or depressive symptoms were conducted. These psychological dimensions are closely linked to stress and could have provided additional insights into the broader mental health effects of the intervention. Additionally, no physiological stress markers (eg, cortisol levels, HRV, and electrodermal skin activity) or other objective parameters were collected. Sole reliance on self-reported stress levels introduces potential biases (eg, social desirability), which may have affected the accuracy of the findings. However, as we wanted to keep the intervention as short as possible, we abstained from using an extensive test battery or setup. Future studies should include a multimodal stress assessment, potentially integrating real-time biometric data using wearable technology or mixed-reality applications. The single-center design and small sample size may also limit the generalizability of the results, as factors specific to the study setting could have influenced outcomes. Selection bias may also have influenced the results, as participants might have been particularly motivated, tech-savvy, or predisposed to respond positively to VR-based interventions. This self-selection could limit the generalizability of the findings to a broader population. The potential for a novelty effect must also be acknowledged. Participants’ stress reduction could partially stem from the excitement or novelty of using VR technology rather than the intervention’s intrinsic therapeutic effects. Furthermore, this study did not assess long-term effects. The sustainability of stress reduction over time remains unclear, and follow-up assessments would be necessary to determine whether the observed benefits persist beyond the immediate post-intervention period.

Ultimately, while the results are encouraging, future research should focus on a randomized controlled design, incorporate a multimodal assessment of stress, depression, or burnout, including objective biological stress markers, assess long-term effects, and involve larger, more diverse populations to strengthen the evidence base for VR interventions in stress management in the health care setting. Furthermore, it is essential to identify the specific aspects of the experience that elicit the most significant responses. For example, archival data before and after the Covid-19 pandemic show that passive content with less interactivity resulted in a greater positive mood state after the COVID-19 onset, likely related to its capacity to reduce stress, facilitate restoration, and improve persistent affective states in stressful environments [39].

Conclusions

In summary, this pilot study adds to the growing evidence supporting the use of VR for workplace well-being by demonstrating the feasibility and short-term effectiveness of immersive VR simulations for stress reduction among emergency physicians. A brief VR-based relaxation break conducted directly in the ED workplace significantly decreased subjective stress levels, with high user satisfaction and minimal side effects reported. However, implementation challenges were evident, primarily due to the significant time constraints faced by health care professionals in this high-pressure environment. These findings highlight the potential of VR as a tool to enhance workplace well-being while underscoring the need for strategies to overcome logistical barriers and better integrate such interventions into routine clinical practice. Future studies should

focus on long-term effects, objective stress measures, and scalable implementation strategies to further validate and optimize this approach.

Acknowledgments

We would like to thank the entire team at the Department of Emergency Medicine at the University Hospital Bern, Switzerland, for their openness to this intervention and their willingness to participate in the study. The generative AI tool ChatGPT4.0 by OpenAI was used for language editing.

Data Availability

Data contain potentially identifying or sensitive employee information. Data used in this study are available upon reasonable request from the corresponding author at the Emergency Department of the University Hospital Bern, Switzerland to researchers eligible under Swiss legislation to work with codified research data.

Authors' Contributions

All authors contributed to the design of the project. TB and TCS conceptualized the study. SH and SS collected the data. SH, TB, and TCS analyzed and interpreted the data. TB and SH wrote the manuscript. All authors revised, reviewed and approved the manuscript before submission. TB and SH contributed equally.

Conflicts of Interest

TCS holds the endowed professorship of emergency telemedicine at the University of Bern sponsored by the Touring Club Switzerland. The sponsor has no influence on the research or decision to publish.
All other authors have nothing to disclose.

References

1. Rotenstein LS, Torre M, Ramos MA, et al. Prevalence of burnout among physicians: a systematic review. JAMA 2018 Sep 18;320(11):1131-1150. [doi: 10.1001/jama.2018.12777] [Medline: 30326495]
2. Stehman CR, Testo Z, Gershaw RS, Kellogg AR. Burnout, drop out, suicide: physician loss in emergency medicine, Part I. West J Emerg Med 2019 May;20(3):485-494. [doi: 10.5811/westjem.2019.4.40970] [Medline: 31123550]
3. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018 Jun;283(6):516-529. [doi: 10.1111/joim.12752] [Medline: 29505159]
4. Vanyo L, Sorge R, Chen A, Lakoff D. Posttraumatic stress disorder in emergency medicine residents. Ann Emerg Med 2017 Dec;70(6):898-903. [doi: 10.1016/j.annemergmed.2017.07.010] [Medline: 28826753]
5. McBeth BD, Ankel FK, Ling LJ, et al. Substance use in emergency medicine training programs. Acad Emerg Med 2008 Jan;15(1):45-53. [doi: 10.1111/j.1553-2712.2007.00008.x] [Medline: 18211313]
6. Schernhammer ES, Colditz GA. Suicide rates among physicians: a quantitative and gender assessment (meta-analysis). Am J Psychiatry 2004 Dec;161(12):2295-2302. [doi: 10.1176/appi.ajp.161.12.2295] [Medline: 15569903]
7. Heymann EP, Romann V, Lim R, et al. Physician wellbeing and burnout in emergency medicine in Switzerland. Swiss Med Wkly 2024 May 16;154(5):3421. [doi: 10.57187/s.3421] [Medline: 38753467]
8. Lu DW, Dresden S, McCloskey C, Branzetti J, Gisondi MA. Impact of burnout on self-reported patient care among emergency physicians. West J Emerg Med 2015 Dec;16(7):996-1001. [doi: 10.5811/westjem.2015.9.27945] [Medline: 26759643]
9. WHO Guidelines on Mental Health at Work, 1st edition: World Health Organization; 2022. URL: https://www.who.int/publications/i/item/9789240053052 [accessed 2025-06-26]
10. Riches S, Taylor L, Jeyarajaguru P, Veling W, Valmaggia L. Virtual reality and immersive technologies to promote workplace wellbeing: a systematic review. J Ment Health 2024 Apr;33(2):253-273. [doi: 10.1080/09638237.2023.2182428] [Medline: 36919828]
11. West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. The Lancet 2016 Nov;388(10057):2272-2281. [doi: 10.1016/S0140-6736(16)31279-X]
12. Pizzoli SFM, Mazzocco K, Triberti S, Monzani D, Alcañiz Raya ML, Pravettoni G. User-centered virtual reality for promoting relaxation: an innovative approach. Front Psychol 2019;10:479. [doi: 10.3389/fpsyg.2019.00479] [Medline: 30914996]
13. Chuan A, Zhou JJ, Hou RM, Stevens CJ, Bogdanovych A. Virtual reality for acute and chronic pain management in adult patients: a narrative review. Anaesthesia 2021 May;76(5):695-704. [doi: 10.1111/anae.15202] [Medline: 32720308]
14. Birrenbach T, Bühlmann F, Exadaktylos AK, Hautz WE, Müller M, Sauter TC. Virtual Reality for pain relief in the emergency room (VIPER) - a prospective, interventional feasibility study. BMC Emerg Med 2022 Jun 21;22(1):113. [doi: 10.1186/s12873-022-00671-z] [Medline: 35729502]

15. Zhang Q, Fu Y, Lu Y, et al. Impact of virtual reality-based therapies on cognition and mental health of stroke patients: systematic review and meta-analysis. J Med Internet Res 2021 Nov 17;23(11):e31007. [doi: 10.2196/31007] [Medline: 34787571]
16. Emmelkamp PMG, Meyerbröker K. Virtual reality therapy in mental health. Annu Rev Clin Psychol 2021 May 7;17(1):495-519. [doi: 10.1146/annurev-clinpsy-081219-115923] [Medline: 33606946]
17. Riches S, Jeyarajaguru P, Taylor L, et al. Virtual reality relaxation for people with mental health conditions: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2023 Jul;58(7):989-1007. [doi: 10.1007/s00127-022-02417-5] [Medline: 36658261]
18. Ladakis I, Filos D, Chouvarda I. Virtual reality environments for stress reduction and management: a scoping review. Virtual Real 2024 Mar;28(1):50. [doi: 10.1007/s10055-024-00943-y]
19. Navarro-Haro MV, López-Del-Hoyo Y, Campos D, et al. Meditation experts try virtual reality mindfulness: a pilot study evaluation of the feasibility and acceptability of virtual reality to facilitate mindfulness practice in people attending a mindfulness conference. PLoS ONE 2017;12(11):e0187777. [doi: 10.1371/journal.pone.0187777] [Medline: 29166665]
20. Riches S, Azevedo L, Bird L, Pisani S, Valmaggia L. Virtual reality relaxation for the general population: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2021 Oct;56(10):1707-1727. [doi: 10.1007/s00127-021-02110-z] [Medline: 34120220]
21. Fauveau V, Filimonov AK, Pyzik R, et al. Comprehensive assessment of physiological and psychological responses to virtual reality experiences. Journal of Medical Extended Reality 2024 Mar 1;1(1):227-241. [doi: 10.1089/jmxr.2024.0020]
22. Mazgelytė E, Rekienė V, Dereškevičiūtė E, et al. Effects of virtual reality-based relaxation techniques on psychological, physiological, and biochemical stress indicators. Healthcare (Basel) 2021 Dec 14;9(12):1729. [doi: 10.3390/healthcare9121729] [Medline: 34946455]
23. Kaplan S. The restorative benefits of nature: toward an integrative framework. J Environ Psychol 1995 Sep;15(3):169-182. [doi: 10.1016/0272-4944(95)90001-2]
24. Pascual K, Fredman A, Naum A, Patil C, Sikka N. Should mindfulness for health care workers go virtual? A mindfulness-based intervention using virtual reality and heart rate variability in the emergency department. Workplace Health Saf 2023 Apr;71(4):188-194. [doi: 10.1177/21650799221123258] [Medline: 36377263]
25. Beverly E, Hommema L, Coates K, et al. A tranquil virtual reality experience to reduce subjective stress among COVID-19 frontline healthcare workers. PLoS ONE 2022;17(2):e0262703. [doi: 10.1371/journal.pone.0262703] [Medline: 35139092]
26. Nijland J, Veling W, Lestestuiver BP, Van Driel CMG. Virtual Reality Relaxation for Reducing Perceived Stress of Intensive Care Nurses During the COVID-19 Pandemic. Front Psychol 2021;12:706527. [doi: 10.3389/fpsyg.2021.706527] [Medline: 34659021]
27. Adhyaru JS, Kemp C. Virtual reality as a tool to promote wellbeing in the workplace. Digit Health 2022;8:20552076221084473. [doi: 10.1177/20552076221084473] [Medline: 35284084]
28. Exadaktylos AK, Hautz WE. Emergency medicine in Switzerland. ICU Manag Pract 2015;15:160-163 [FREE Full text]
29. Lesage FX, Berjot S. Validity of occupational stress assessment using a visual analogue scale. Occup Med 2011 Sep 1;61(6):434-436. [doi: 10.1093/occmed/kqr037]
30. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
31. Usoh M, Catena E, Arman S, Slater M. Using presence questionnaires in reality. 2006. URL: https://direct.mit.edu/pvar/article-abstract/9/5/497/18368/Using-Presence-Questionnaires-in-Reality?redirectedFrom=fulltext [accessed 2025-06-26]
32. Putrino D, Ripp J, Herrera JE, et al. Multisensory, nature-inspired recharge rooms yield short-term reductions in perceived stress among frontline healthcare workers. Front Psychol 2020;11:560833. [doi: 10.3389/fpsyg.2020.560833] [Medline: 33329188]
33. Gaggioli A, Pallavicini F, Morganti L, et al. Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. J Med Internet Res 2014 Jul 8;16(7):e167. [doi: 10.2196/jmir.3235] [Medline: 25004803]
34. Nicksic Sigmon CA, Park SY, Bam D, Rimel SE, Boeldt DL. Feasibility and acceptability of virtual reality mindfulness in residents, fellows, and students at a university medical campus. Mayo Clin Proc Digit Health 2023 Dec;1(4):467-475. [doi: 10.1016/j.mcpdig.2023.07.006] [Medline: 40206320]
35. Linares-Chamorro M, Domènech-Oller N, Jerez-Roig J, Piqué-Buisan J. Immersive therapy for improving anxiety in health professionals of a regional hospital during the COVID-19 pandemic: a quasi-experimental pilot study. Int J Environ Res Public Health 2022 Aug 9;19(16):9793. [doi: 10.3390/ijerph19169793] [Medline: 36011433]
36. Ferrer Costa J, Moran N, Garcia Marti C, Colmenares Hernandez LJ, Radu Ciorba Ciorba F, Ciudad MJ. Immediate impact of an 8-week virtual reality educational program on burnout and work engagement among health care professionals: pre-post pilot study. JMIR XR Spatial Comput 2024;1:e55678. [doi: 10.2196/55678]
37. Schrempf MC, Petzold J, Petersen MA, et al. A randomised pilot trial of virtual reality-based relaxation for enhancement of perioperative well-being, mood and quality of life. Sci Rep 2022 Jul 14;12(1):12067. [doi: 10.1038/s41598-022-16270-8] [Medline: 35835944]

38. Beckham A, Cooper N. Well-being interventions for emergency department staff: ‘necessary’ but ‘inadequate’ – a phenomenographic study. Emerg Med J 2025 Jan;42(1):9-13. [doi: 10.1136/emermed-2023-213852]
39. Housand J, Cornelius A, Shackleford KE. Greater usage and positive mood change for users of a dynamic VR app before and after the COVID-19 pandemic onset. Front Psychol 2024;15:1278207. [doi: 10.3389/fpsyg.2024.1278207] [Medline: 38476392]

Abbreviations

ED: emergency department
IQR: interquartile range
NRS: numeric rating scale
PSS-14 : Perceived Stress Scale 14
SSQ: Simulator Sickness Questionnaire
VR: virtual reality

Edited by L Riedemann; submitted 13.02.25; peer-reviewed by J Ogunsakin, M Gasmi, N Mungoli; revised version received 27.03.25; accepted 04.04.25; published 04.07.25.

Please cite as:
Birrenbach T, Häni S, Jegerlehner S, Schober S, Exadaktylos AK, Sauter TC
Feasibility, Subjective Effectiveness, and Acceptance of Short Virtual Reality Relaxation Breaks for Immediate Perceived Stress Reduction in Emergency Physicians: Single-Arm Pre-Post Intervention Study
JMIR XR Spatial Comput 2025;2:e72605
URL: https://xr.jmir.org/2025/1/e72605
doi:10.2196/72605

The Feasibility and Acceptability of a Stand-Alone Virtual Reality Headset on Perceived Pain and Anxiety During Bone Marrow Biopsies: Mixed Methods Pilot Study

Ajay Mittal[1,2], DO; Suhaiba Huq[1], BS; Jonathan Wakim[3], MD, MBA; Kevin Kapadia[4], MA; Tung Wynn[1], MD

[1]Department of Medicine, University of Florida College of Medicine, 1600 SW Archer Rd, Gainesville, FL, United States

[2]BayCare Health System/Northwestern University Feinberg School of Medicine, Clearwater, FL, United States

[3]Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

[4]Department of Medicine, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Suhaiba Huq, BS
Department of Medicine, University of Florida College of Medicine, 1600 SW Archer Rd, Gainesville, FL, United States

Abstract

Background: Virtual reality (VR) is an emerging technology that provides an immersive user experience and has the ability to distract patients from the negative or painful experiences commonly associated with medical procedures. Bone marrow biopsies are medical procedures where a needle is inserted into the bone and a syringe is used to withdraw the liquid bone marrow. They are performed to diagnose and monitor disorders affecting the blood, often as part of care for hematology and patients with cancer.

Objective: The purpose of this pilot study is to assess the feasibility of VR as an adjunctive therapy to alleviate the perception of pain and anxiety in patients receiving bone marrow biopsies.

Methods: This pilot study enrolled 60 adult participants receiving a bone marrow biopsy to assess the acceptability and feasibility of VR to impact reported pain and anxiety levels compared to the participants' baseline measurements preoperatively. They were randomly assigned to "control"/non-VR intervention (n=30) and "experimental"/VR groups (n=30). The "experimental"/VR group used the Meta Quest 2 headset (Meta) with original VR content developed for this study. Participants completed a survey adapted from a standardized verbal numerical rating scale to rate their pain and anxiety levels before and after the bone marrow biopsy. Measurements such as procedure length, patient vitals, and experience were also gathered from both study groups.

Results: Results indicated that participants had no significant differences in their heart rate, respiration rate, and blood oxygen saturation levels between the 2 groups. Participants in the VR group (n=30) had a significantly shorter procedure length than the control group (n=30) with a 25% time reduction (P=.02). Participants in the VR group (mean 4.29, SD 1.19) were significantly more likely to rate the distraction as effective (P<.001) and report they would repeat the procedure (mean 4.32, SD 1.05; P<.001). Finally, participants in the VR group (mean 2.13, SD 1.26) had significantly lower levels of anxiety during the procedure (P<.001) and felt significantly more comfortable after the procedure (mean 4.45, SD 1.12; P<.001).

Conclusions: This investigation encourages the acceptability of using VR intervention for patients undergoing bone marrow biopsies. Further, the length of procedures was found to be shorter when compared to the control group, supporting the feasibility of the technology for clinical management. These novel interventions can provide distraction-based therapy that is noninferior to the standard of care and provide enjoyable user experiences that reduce the perceived pain and anxiety of nonsedated medical procedures.

(JMIR XR Spatial Comput 2025;2:e65324) doi:10.2196/65324

KEYWORDS

virtual reality; bone marrow biopsy; oncology; pharmacologic pain management; pain management

Introduction

Standard pain management protocols in adult medicine settings, hospitals, or clinics, typically rely on the use of pharmacotherapies such as acetaminophen, nonsteroidal anti-inflammatory drugs, analgesics, and opioids to alleviate acute pain [1]. Although these forms of treatment can be effective, there are growing concerns surrounding the potential health risks associated with the use of certain pharmacotherapies, such as opioid addiction, across patient populations [2]. As a result, there is a growing interest in reducing the use of pharmacotherapy for the treatment of acute pain in favor of nonpharmacologic options [3]. However, there are few

alternative options for providing nonpharmacologic therapy in adults, which often results in inadequate pain control [4]. There is a need to find feasible and acceptable adjunct forms of pain management and anxiety reduction in a hospital or clinic setting.

Distraction techniques used by hospital staff help patients cope with injuries, hospitalization, or illness and differ based on each patient's needs and preferences [5]. Some common techniques used to distract from chronic pain include controlled breathing, guided imagery, and relaxation [6]. Passive techniques such as auditory distraction and television have been used to distract from acute pain resulting from routine procedures [7]. Studies for each of these different techniques over the years found some positive, but mostly mixed results [8]. Thus, there is still limited evidence demonstrating the effectiveness of distraction [9]. Moreover, there is no conclusive study suggesting that one technique supersedes others since each patient has different preferences, medical situations, behavioral needs, and developmental needs.

Virtual reality (VR) is an emerging technology that provides an immersive user experience [10]. These immersive VR experiences show promise as a tool that can reduce perceived pain and anxiety related to acute pain in emergency rooms or other clinical settings by lessening the vividness of memories [11]. One study has even observed that immersive VR was more effective than standard care in ameliorating pain among pediatric patients undergoing venipunctures [12]. It is theorized that immersive VR draws the patient's attention away from aversive, or painful, stimuli by keeping their focus on something more engaging [13]. A study published in 2023 by Alaniz et al [14] found VR to be an underused intraoperative tool that enhances the overall patient experience in the emergency room. Furthermore, another study by Sabinash [15] reported positive findings with this technology, which are tempered by the limitations of current research, but VR still holds great potential. The stand-alone headset's hardware offers an affordable alternative that can potentially provide a nonpharmacologic treatment for acute pain management and anxiety reduction [16]. A study published in Virtual Worlds found that it is likely to see the widespread implementation of VR in health care in the coming decade [17]. With recent advancements, VR technology is used in numerous settings, but limited research has been done thus far to show how effective it is [18].

The purpose of this pilot study is to evaluate the feasibility and acceptability of using VR adjunctively as a nonpharmacological distraction method for bone marrow biopsies to reduce perceived anxiety and pain in an oncology ward setting. The information gained will help to design future studies needed to ensure the efficacy and reliability of the devices.

Methods

Design Overview

This pilot acceptability and feasibility study was performed at the University of Florida (UF) Shands Cancer Hospital to assess VR as a relaxation and distraction tool for patients receiving a bone marrow biopsy procedure. We used a mixed methods design, in which trained research assistants identified and determined eligibility of patients admitted to the UF Health Division of Hematology & Oncology Ward in coordination with medical staff. Participants were randomly assigned to the control group and experimental group through a random number sequence to determine group allocation. No stratification was used in this process. The control (or non-VR intervention) group did not receive any special distraction in addition to local anesthesia, following conventional clinical practice for managing discomfort. Local anesthesia is defined as 5‐10 mL of lidocaine 1% (10 mg/mL) without epinephrine administered to patients for bone marrow biopsies, according to the UF Health Shands Hospital protocol. Approximately 1‐2 mL skin wheal (intradermal) was used to numb the skin, followed by a 2‐3 mL subcutaneous infiltration to numb deeper tissue, and then a 2‐5 mL periosteal infiltration for the bone surface. No additional nonpharmacological distraction methods were implemented in this group. Data regarding patient vitals, procedure length, patient feelings before and after the procedure, and patient experiences with the procedure were compared between participants assigned to the VR and control groups. The 2 groups were compared using independent *t* tests with Cohen *d* measure of effect size. For any measure taken before and after the procedure, the difference between the 2 was calculated and used to compare the groups for the *t* tests.

Training Protocol

Research assistants underwent a 4‐ to 5-week training period covering all aspects of the patient intervention, including preprocedural preparation, informed consent, intervention administration, and postprocedural procedures. Training emphasized appropriate patient interaction, including how to explain and review consent and procedural forms. In addition, research assistants received instruction on infection control measures, such as proper sanitization of the VR equipment. These measures for the VR headsets included the use of Super Sani-Cloth Germicidal Disposable Wipes to thoroughly disinfect the headset and controllers before and after each patient's use. In addition, research assistants were instructed to wash their hands before and after each patient interaction to minimize the risk of cross-contamination. The final 2 weeks of training focused on VR equipment operation, troubleshooting technical issues, and ensuring a smooth patient experience.

Participants

Participants were recruited between February 2021 and February 2024 from the UF Health Oncology & Hematology ward. A total of 93 patients were initially screened for eligibility, of whom 60 patients met the inclusion criteria and were enrolled in the study, with 30 randomized to the VR group and 30 to the control group. Participants were included according to the following criteria: have a scheduled bone marrow biopsy, aged more than 18 years, and be able to physically wear and tolerate the VR headset. Participants were excluded if they had nausea or vomiting upon admission, required urgent procedures or were otherwise deemed unstable by hospital staff, had a condition that prevents the use of VR technology such as epilepsy, or a facial or scalp wound, had any visual, hearing, or cognitive impairments that would limit their ability to take part in the study, or if they could not read, speak, or write in English. The

demographic characteristics of the participants are shown in Table 1. The majority of participants were male, the median age was 61.5 years (IQR 53 - 68), the racial composition was 75% White with approximately 17% Black (10/60 participants) and 8% Latino (5/60 participants) enrolled, and 63% (38/60 participants) of them reported some familiarity with VR.

Table . Demographic information and familiarity with VR[a].

Demographics	Total population (N=60)	Control (n=30)	VR (n=30)	Statistical test results for differences between control and VR groups, *t* test or chi-square (*df*)	*P* value
Age (years), median (IQR)	61.5 (53-68)	65 (58.5 - 69.5)	58 (48-64)	1.31 (57.18)[b]	.92
Gender, n (%)				0.16 (1)[c]	.69
Male	36 (60)	16 (53)	19 (63)		
Female	24 (40)	14 (47)	11 (37)		
Race, n (%)				1.86(4)[c]	.76
White	45 (75)	23 (77)	22 (73)		
Black	10 (17)	4 (13)	5 (17)		
Latinx	5 (8)	3 (10)	3 (10)		
VR familiarity, n (%)	38 (63)	19 (63)	19 (63)	0.78 (1)[c]	.78

[a]VR: virtual reality.

[b]*t* test.

[c]Chi-square test.

Distraction-Based VR Intervention

The “HealthPointXR” app (Gainesville, Florida), developed for this study, was used to provide distraction and relaxation for participants enrolled in the experimental group. The “HealthPointXR” app was developed by Drew Gill as part of a UF undergraduate engineering organization by the name of Dream Team Engineering. There are no financial disclosures related to the development or commercialization of this app at this time. The HealthPointXR game is currently an unlicensed, private project owned by Dream Team Engineering as per GitHub for an unlicensed project “the default copyright laws apply, meaning that [the owner] retain[s] all rights to [their] source code and no one may reproduce, distribute, or create derivative works from [their] work.” The code is currently neither open source nor free, but contacting Dream Team Engineering via their website is encouraged for any usage. This study used a retrofitted Meta Quest 2 headset (Meta) with plastic straps and easily sanitizable materials. The VR device was selected based on affordability and quality to ensure its practicality and scalability in clinical settings. The Meta Quest 2 met these criteria, while the Apple Vision Pro was considered as a potential alternative. However, due to its recent market entry and significantly higher cost, it was not incorporated into this study. The Meta Quest 2 fits most head shapes and sizes. Tablets were connected to the headsets for research assistants to control and supervise the VR content in real time using an intuitive interface. The tablet and the VR headset were connected through a Bluetooth pairing and did not require internet or Wi-Fi connectivity. “HealthPointXR” VR game included an experience of a tranquil walking path through a natural environment where the user snaps photographs of wildlife passing by while being transported in a cart along a track. The user must take these different photographs when prompted to advance their journey. Users are also incentivized to follow the track closely through minimal head movements in order to receive points on the trip across the natural environment to further immerse the user. Soothing music and different sounding notifications provide auditory cues to engage the user.

For participants randomly selected to the experimental group, research assistants provided a standardized 5-minute tutorial on wearing the VR headset and fitting it properly for optimal user experience. The length of time required to fit the VR headset on participants was not factored into the total length of the procedure, as it was done during normally occurring time intervals prior to the bone marrow biopsy procedure. Figure 1 shows an example of the VR experience.

Figure 1. "HealthPointXR" virtual reality experience.

Patient Vitals and Procedure

Patients in both the VR and the control group had their vitals, including heart rate, respiration rate, and blood oxygen saturation, recorded through a pulse oximeter during and after the procedure. These vitals were taken by the health care staff present at the time of the procedure, and research assistants collected these data from the health care staff after completion of the intervention. Procedure length was recorded on a stopwatch by the research assistants, who began measuring the procedure length as soon as the VR headset was placed on the patients. Timing concluded at the end of the medical procedure once the clinician completed the last component of the procedure.

Patient-Reported Emotional and Procedural Experience Measures

Patients in both the VR and the control group were asked the same questions regarding their feelings and experience with the procedure. Patients' feelings regarding the procedure were assessed by pain before and after the procedure, worry before and after the procedure, anxiety before and after the procedure, and comfort after the procedure. Feelings of pain were recorded using a verbal numerical rating scale from 1 to 10 before and after the procedure with higher values indicating higher feelings of pain. Feelings of worry were measured before and after the procedure using a scale from 1 to 5, with 1 indicating "not worried at all" and 5 indicating "extremely worried." Feelings of anxiety during the procedure were recorded on a verbal numerical rating scale from 1 to 5, with higher values indicating higher feelings of anxiety. Feelings of comfort after the procedure were evaluated using a 5-point scale from strongly disagree to strongly agree for the statement "I felt comfortable with the distraction."

Patients' experiences with the procedure were all assessed on a 5-point scale from strongly disagree to strongly agree. To measure awareness, patients were asked their agreement with the following statement: "I was very aware of the procedure I was receiving." To measure distraction effectiveness, patients were asked their agreement with "Using the distraction made me feel less worried about getting the procedure." Finally, for likelihood to repeat, patients were asked their agreement with "I would want to use the same distraction on medical procedures in the future."

Ethical Considerations

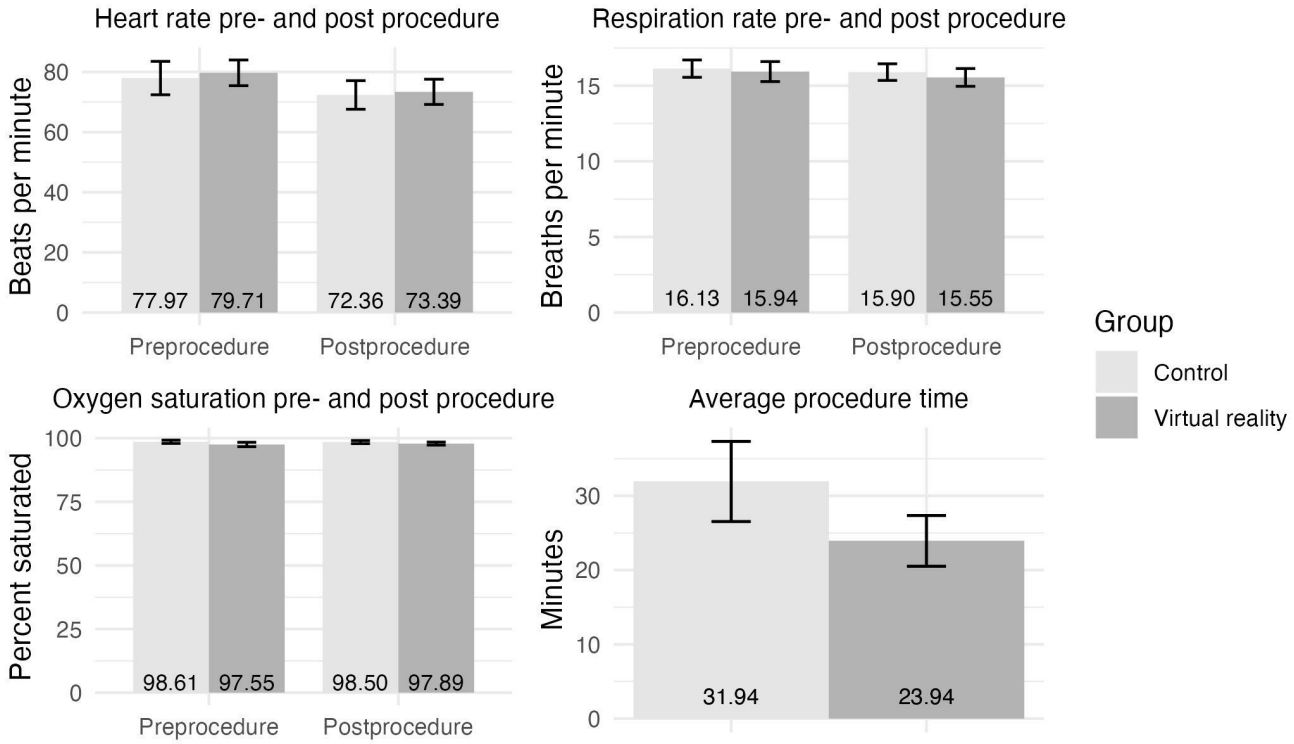
This study, which includes human participant research, was approved by the Institutional Review Board of the Florida Department of Health (IRB201900850). The informed consent forms, used in this study, provided participants with a description of the study, its qualitative and quantitative measures, potential risks and discomforts, and explicitly asked the patient to voluntarily agree to participate and allow their data to be collected. Study data are deidentified as all participants were assigned a numerical code. This code follows the format of OC001. Anonymity of all study subjects is ensured. No compensation of any sort was offered to human participants. This was stated in the informed consent form. No identification of individual participants is present in any images of this study or in any supplemental material.

Results

Patient and Procedure Summary Information

Figure 2 summarizes the patient's vitals and procedure duration between the VR and the control group. The patient's heart rate, respiration rate, and blood oxygen saturation were taken before and after the procedure. There were no significant differences in the change in vitals pre- and postprocedure between the VR and control group. The average procedure duration was significantly shorter in the VR group (mean 23.94, SD 9.72) compared to the control group (mean 31.94, SD 15.34; t_{51}=−2.45, P=.02, d=0.62).

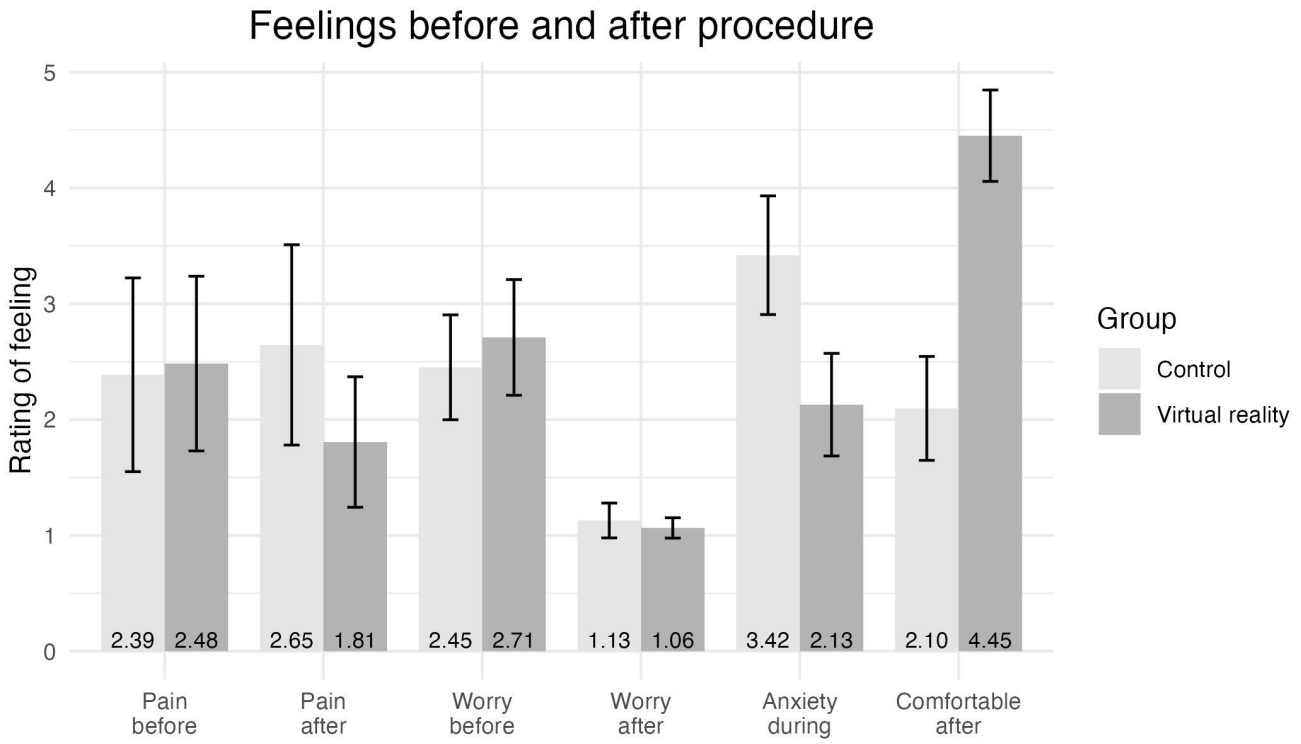
Figure 2. Summary of patient vitals and procedure length.



Patient-Reported Emotional Measures

Figure 3 summarizes patients' feelings before, during, and after the procedure between the VR and the control group. Patients were asked to indicate their feelings as explained in the Methods section, with higher scores indicating higher levels of pain, worry, anxiety, and comfort. The error bars represent 95% CIs, and the mean for each group is labeled at the bottom. Participants in the VR group (mean 2.13, SD 1.26) had significantly lower feelings of anxiety during the procedure compared to the control group (mean 3.42, SD 1.46; t_{57}=−3.73, P<.001, d=0.95). Participants in the VR group (mean 4.45, SD 1.12) had significantly higher feelings of comfort with the distraction compared to the control group (mean 2.10, SD 1.27; t_{59}=7.73, P<.001, d=1.96). There were no significant differences between the VR and control groups for pain and worry either before or after the procedure. In addition, there was no significant difference within either the VR or control group for the difference in pain or worry perception before and after the procedure.

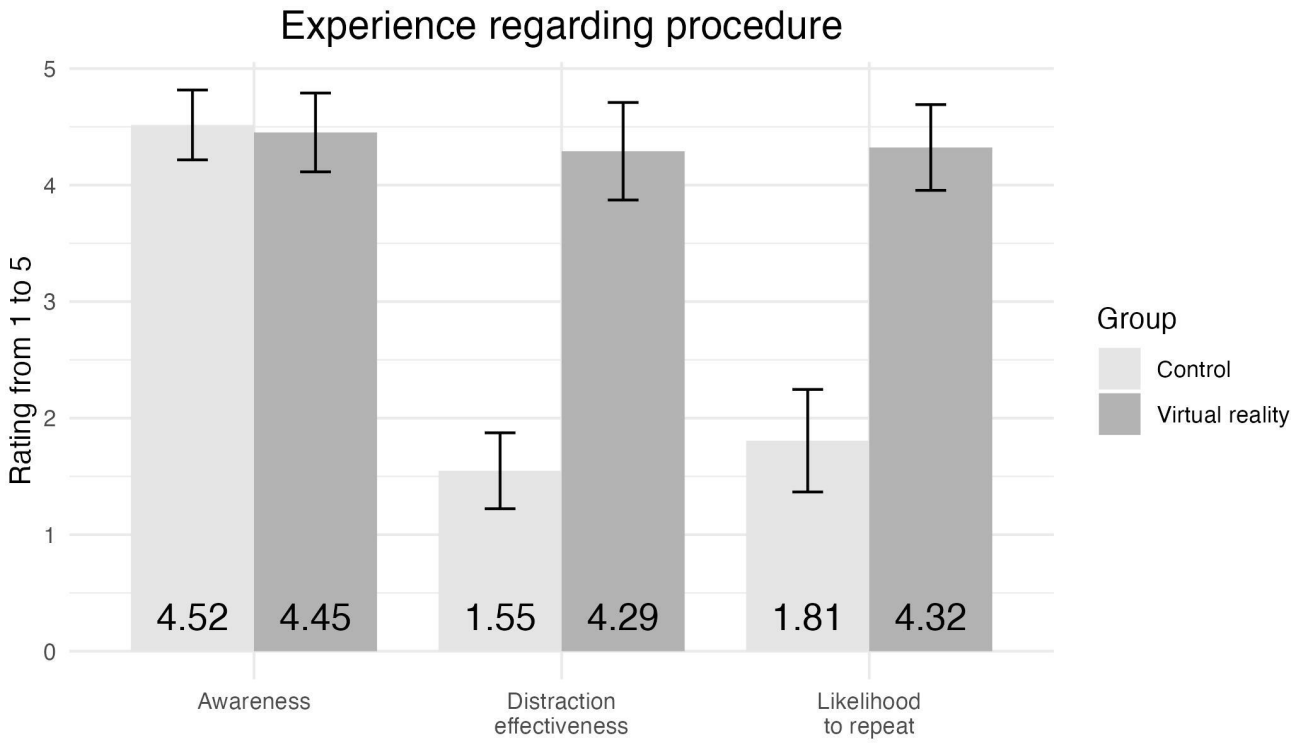
Figure 3. Differences in patient feelings before and after the procedure.



Patient-Reported Procedural Experience Measures

Figure 4 summarizes differences in patient experience during the procedure between the VR and the control group. Patients were asked to rate the statements on a scale of 1 to 5, with higher scores indicating a larger awareness of the procedure occurring, higher effectiveness of the distraction, and a greater likelihood to repeat the procedure. The error bars represent 95% CIs and the mean for each group is labeled at the bottom. Participants in the VR group (mean 4.29, SD 1.19) were significantly more likely compared to the control group (mean 1.55, SD 0.93) to rate the distraction as effective (t_{57}=10.14, P<.001, d=2.57). Participants in the VR group (mean 4.32, SD 1.05) were also significantly more likely compared to the control group (mean 1.81, SD 1.25) to indicate they would repeat the experience (t_{58}=8.60, P<.001, d=2.18). However, there were no significant differences between participants in the VR group (mean 4.45, SD 0.96) and the control group (mean 4.52, SD 0.85) of how aware participants were of the procedure occurring (t_{57}=−0.28, P=.78, d=0.07).

Figure 4. Differences in patient experience during the procedure.



Power Analysis

We conducted a post hoc power analysis using the *pwr* package in R to assess the statistical power of our independent samples *t* tests. Using the calculated effect sizes, we estimated the achieved power for each test assuming a 2-tailed *t* test with an alpha level of .05 and a sample size of 30 per group. The resulting power values ranged from 0.058 (d=0.07) to 1 (d=2.57). The estimated achieved power for 4 of the 6 independent samples *t* tests was above 0.95 (d≤0.95).

Discussion

Principal Findings

The results of this pilot study suggest that VR has a role in reducing anxiety and perceived pain in adult patients receiving bone marrow biopsy. Although this study is not designed to be statistically powered, a notable reduction in perceived pain was measured in the VR group, with a reduction of 0.7 in reported pain levels postoperatively compared to the control group's increase of 0.2 reported pain postoperatively on a 10-point scale. These findings paralleled those reported by Wong and Choi [12] when assessing patient perceptions of VR for pain relief in labor [19].

Anxiety levels before and after bone marrow biopsies were not significantly different in the VR group compared to the control group. Anxiety in patients with cancer can be triggered by a multitude of factors, as well as anticipation of pain associated with medical procedures [20]. VR might be slightly effective in reducing reported anxiety levels, although the complex combination of factors associated with the care of patients with cancer requires caution when interpreting the results of a tool used in an acute care application. This differs from similar studies that have reported marked changes in reported anxiety levels. A study by Fabi [21] reported that among 22 patients who used VR during chemotherapy, there was a significant decrease in perceived anxiety and duration of procedure compared to the control group.

Acceptability of VR in the hospital setting was supported by the patient experience findings and reinforced by the anxiety and perceived pain data collected during this study. The results of the surveys suggest a 2 - 3 point higher score in level of immersion, likelihood to use intervention again, and comfort level in the VR group compared to the control group using a 5-point scale. The majority of patients expressed a willingness to use VR as a distraction-based tool during their bone marrow biopsy. Furthermore, a study published by the *Journal of Internet Medical Research* in 2022 found VR beneficial for breaking up the monotony of treatment, providing an additional choice of activity, and in some instances a distraction from the treatment itself [22].

Measurements of feasibility for VR focused on the usability and comfort of the technology as well as its impact on the length of time for the medical procedure to be completed. To maintain methodological rigor, confounding variables were controlled as much as possible to ensure that any observed differences could be attributed primarily to the VR intervention rather than external factors. These confounding factors include standardizing the administration of lidocaine to 5 - 10 mL for bone marrow biopsies within the hospital's protocol, having the same clinician perform all the bone marrow biopsies throughout the course of this investigation, timing the procedure length to reflect the total length of the procedure as soon as a distraction method begins, that is when a VR program is turned on, before the procedure, as well as maintaining consistent informed consent and randomization practices for all participants. Consequently, the time differences observed in the

results are attributed to the impact of VR. Notably, there was an 8-minute faster completion or 25% time reduction for completion of the bone marrow biopsy procedure among the VR intervention group (n=30) compared to the control group (n=30). The faster completion of the procedure occurred without any report of compromised quality of care, and these data provide a valuable clinical benchmark for the use of VR in the hospital setting, as it appears to improve the efficiency of providers while performing routine procedures. This surprising difference in procedure length is mostly attributed to the patients in the VR group being more composed and less tense at the beginning of the procedure, allowing for clinicians to more efficiently administer local anesthetic and perform the biopsy in a shorter amount of time. As the same clinician performed all of these bone marrow biopsies, clinician-to-clinician skill level was not a factor modulating the findings. The majority of the time was saved with the patient being immersed in the virtual world at the beginning of the procedure, which coincided with lower self-reported anxiety and pain ratings. It is important to note that the fitting of the VR headset on participants was performed during usual preparatory intervals prior to the biopsy procedure, and thus was not included in the recorded procedure time. The same size remains a limitation on these findings to differentiate them from being anecdotal as opposed to being powered, warranting the need for further investigation regarding time efficiencies and cost-effectiveness regarding the use of VR in minimally invasive procedures such as bone marrow biopsies.

The existing evidence to support the implementation of distraction-based VR therapy is still limited among patients with cancer receiving bone marrow biopsies [23]. This study provided insight into the novel application of a stand-alone VR headset. The results from this study indicate the use of VR as a noninferior adjunct tool that is acceptable and feasible for providing distraction and relaxation in adult patient populations undergoing bone marrow biopsy. Furthermore, these data add to building the evidence base for VR in medicine as part of innovative clinical practice involving digital therapeutics.

This study has several limitations, as the study was not designed to measure the impact of potential moderators on the outcomes. For example, the data cannot provide reliable information on the potential impact of age, gender, or pain tolerance of participants enrolled in the study. Another limitation of this study is the potential for selection bias, as participants were recruited exclusively from the UF Shands Oncology Ward in Gainesville, Florida. In addition, only patients undergoing the specific procedure of bone marrow biopsies were included, which may limit the generalizability of the findings to broader oncology populations or patients undergoing different procedures. The demographic and geographic constraints of the sample may also impact external validity, as factors such as regional health care access, socioeconomic status, and cultural background were not accounted for. The study design, sample size, and inability to double-blind the participants are also notable limitations that impacted the quality of data reported in this research. A review by Kouijzer et al [24] supports these considerations, as a lack of time and expertise on how to use VR in treatment, a lack of personalization of some VR apps to patient needs and treatment goals, or the gap in knowledge on the added value of VR in a specific setting were noted as limitations. In addition, the enrollment strategy could not control for the use of additional pain medication prior to the bone marrow biopsy that was self-prescribed by the patient prior to being approached regarding participation in the study. Furthermore, the study did not factor in medications that the patients are on or the full past medical history of participants to consider the impact of those factors on the participants' vital signs.

Conclusions

In conclusion, this pilot study's findings suggest that the use of an Oculus Quest 2 Headset with the "HealthPointXR" VR program can serve as a noninferior adjunct distraction-based therapy option for patients who are distressed when undergoing a bone marrow biopsy. Based on the parameters assessed, it is reasonable to determine that the implementation of VR improved anxiety levels and perceived pain and did not prolong clinical workflow on average. Further studies with greater statistical power and a larger scale could provide more robust evidence to support the routine application of this emerging technology to impact patients' experience during nonsedated, minimally invasive procedures.

Acknowledgments

We appreciate the support of Drew Gill, Carey Hudson, and the research assistants of Equal Access Research in advancing this study.

Authors' Contributions

AM and JW conceptualized and planned out the methodology of the study. AM and JW worked on the programming and software development of the virtual reality game. TW supervised the project. AM and SH carried out the investigation, and AM, SH, and KK wrote and edited the manuscript. KK analyzed the data, and AM and JW contributed to the interpretation of the results. All authors provided critical feedback throughout the study and helped shape the research and manuscript.

Conflicts of Interest

None declared.

References

1. Hyland SJ, Wetshtein AM, Grable SJ, Jackson MP. Acute pain management pearls: a focused review for the hospital clinician. Healthcare (Basel) 2022 Dec 22;11(1):34. [doi: 10.3390/healthcare11010034] [Medline: 36611494]
2. Phillips JK, Ford MA, Bonnie RJ, et al. Pain management and the intersection of pain and opioid use disorder. In: Pain Management and the Opioid Epidemic: Balancing Societal and Individual Benefits and Risks of Prescription Opioid Use: National Academies Press; 2017. [Medline: 29023083]
3. Demir Y. Non-pharmacological therapies in pain management. In: Pain Management-Current Issues and Opinions: IntechOpen; 2012. [doi: 10.5772/30050]
4. Sikka N, Shu L, Ritchie B, Amdur RL, Pourmand A. Virtual reality-assisted pain, anxiety, and anger management in the emergency department. Telemed J E Health 2019 Dec;25(12):1207-1215. [doi: 10.1089/tmj.2018.0273] [Medline: 30785860]
5. Ibitoye BM, Oyewale TM, Olubiyi KS, et al. The use of distraction as a pain management technique among nurses in a North-central city in Nigeria. Int J Afr Nurs Sci 2019;11:100158. [doi: 10.1016/j.ijans.2019.100158]
6. Vambheim SM, Kyllo TM, Hegland S, Bystad M. Relaxation techniques as an intervention for chronic pain: a systematic review of randomized controlled trials. Heliyon 2021 Aug;7(8):e07837. [doi: 10.1016/j.heliyon.2021.e07837] [Medline: 34485731]
7. Hu W, Yang K, Zhang L, Lu X. Effect of media distraction (audio-visual and music) for pain and anxiety control in patients undergoing shock-wave lithotripsy: a systematic review and meta-analysis. Exp Ther Med 2021 Jun;21(6):623. [doi: 10.3892/etm.2021.10055] [Medline: 33936280]
8. Koller D, Goldman RD. Distraction techniques for children undergoing procedures: a critical review of pediatric research. J Pediatr Nurs 2012 Dec;27(6):652-681. [doi: 10.1016/j.pedn.2011.08.001] [Medline: 21925588]
9. Bukola IM, Paula D. The effectiveness of distraction as procedural pain management technique in pediatric oncology patients: a meta-analysis and systematic review. J Pain Symptom Manage 2017 Oct;54(4):589-600. [doi: 10.1016/j.jpainsymman.2017.07.006] [Medline: 28712986]
10. Rubio-Tamayo JL, Gertrudix Barrio M, García García F. Immersive environments and virtual reality: systematic review and advances in communication, interaction and simulation. Multimodal Technol Interact 2017;1(4):21. [doi: 10.3390/mti1040021]
11. Glennon C, McElroy SF, Connelly LM, et al. Use of virtual reality to distract from pain and anxiety. Oncol Nurs Forum 2018 Jul 2;45(4):545-552. [doi: 10.1188/18.ONF.545-552] [Medline: 29947355]
12. Wong CL, Choi KC. Effects of an immersive virtual reality intervention on pain and anxiety among pediatric patients undergoing venipuncture: a randomized clinical trial. JAMA Netw Open 2023 Feb 1;6(2):e230001. [doi: 10.1001/jamanetworkopen.2023.0001] [Medline: 36795410]
13. Joo Y, Kim EK, Song HG, Jung H, Park H, Moon JY. Effectiveness of virtual reality immersion on procedure-related pain and anxiety in outpatient pain clinic: an exploratory randomized controlled trial. Korean J Pain 2021 Jul 1;34(3):304-314. [doi: 10.3344/kjp.2021.34.3.304] [Medline: 34193636]
14. Alaniz L, Sayadi L, Pakvasa M, et al. Use of virtual reality in emergency room hand procedures. Plast Reconstr Surg Glob Open 2023;11(10S):65-66. [doi: 10.1097/01.GOX.0000992212.01357.a1]
15. Sabinash C. Utilizing virtual reality for distraction in emergency care. Curr Emerg Hosp Med Rep 2023;11(4):143-150. [doi: 10.1007/s40138-023-00280-5]
16. Teh JJ, Pascoe DJ, Hafeji S, et al. Efficacy of virtual reality for pain relief in medical procedures: a systematic review and meta-analysis. BMC Med 2024 Feb 14;22(1):64. [doi: 10.1186/s12916-024-03266-6] [Medline: 38355563]
17. Suh I, McKinney T, Siu KC. Current perspective of metaverse application in medical education, research and patient care. Virtual Worlds 2023;2(2):115-128. [doi: 10.3390/virtualworlds2020007]
18. Smith V, Warty RR, Sursas JA, et al. The effectiveness of virtual reality in managing acute pain and anxiety for medical inpatients: systematic review. J Med Internet Res 2020 Nov 2;22(11):e17980. [doi: 10.2196/17980] [Medline: 33136055]
19. Wong MS, Gregory KD, Spiegel BMR, Khalil C. Patient perceptions of virtual reality for pain relief in labor: a qualitative study. Front Pain Res (Lausanne) 2022;3:1063751. [doi: 10.3389/fpain.2022.1063751] [Medline: 36582195]
20. Sakamoto R, Koyama A. Effective therapy against severe anxiety caused by cancer: a case report and review of the literature. Cureus 2020 Jun 2;12(6):e8414. [doi: 10.7759/cureus.8414] [Medline: 32626629]
21. Fabi A, Fotia L, Giuseppini F, et al. The immersive experience of virtual reality during chemotherapy in patients with early breast and ovarian cancers: the patient's dream study. Front Oncol 2022;12:960387. [doi: 10.3389/fonc.2022.960387] [Medline: 36249001]
22. Janssen A, Fletcher J, Keep M, et al. Experiences of patients undergoing chemotherapy with virtual reality: mixed methods feasibility study. JMIR Serious Games 2022 Feb 21;10(1):e29579. [doi: 10.2196/29579] [Medline: 35188474]
23. Erdős S, Horváth K. The impact of virtual reality (VR) on psychological and physiological variables in children receiving chemotherapy: a pilot cross-over study. Integr Cancer Ther 2023;22:15347354231168984. [doi: 10.1177/15347354231168984] [Medline: 37232275]
24. Kouijzer MMTE, Kip H, Bouman YHA, Kelders SM. Implementation of virtual reality in healthcare: a scoping review on the implementation process of virtual reality in various healthcare settings. Implement Sci Commun 2023 Jun 16;4(1):67. [doi: 10.1186/s43058-023-00442-2] [Medline: 37328858]

Abbreviations

UF: University of Florida
VR: virtual reality

Edited by L Riedemann; submitted 12.08.24; peer-reviewed by M Pritchard, O Liran; revised version received 03.06.25; accepted 06.06.25; published 08.09.25.

Please cite as:
Mittal A, Huq S, Wakim J, Kapadia K, Wynn T
The Feasibility and Acceptability of a Stand-Alone Virtual Reality Headset on Perceived Pain and Anxiety During Bone Marrow Biopsies: Mixed Methods Pilot Study
JMIR XR Spatial Comput 2025;2:e65324
URL: https://xr.jmir.org/2025/1/e65324
doi:10.2196/65324

Augmented Reality in Enhancing Operating Room Crisis Checklist Adherence: Randomized Comparative Efficacy Study

Rayan Ebnali Harari[1*], PhD; Abdullah Altaweel[1,2*], MD; Erik Anderson[3*], MD; Charles Pozner[4*], MD; Rafael Grossmann[5*], MD; Andrew Goldsmith[3], MBA, MD; Hamid Shokoohi[3*], MD

[1]STRATUS Center for Medical Simulation, Mass General Brigham, Harvard Medical School, 10 Vining, Boston, MA, United States

[2]Kuwait Ministry of Health, Kuwait City, Kuwait

[3]Department of Emergency Medicine, Mass General Brigham, Boston, MA, United States

[4]PC Institute for Medical Education, Boston, MA, United States

[5]Department of Surgery, Portsmouth Regional Hospital, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Rayan Ebnali Harari, PhD
STRATUS Center for Medical Simulation, Mass General Brigham, Harvard Medical School, 10 Vining, Boston, MA, United States

Abstract

Background: Effective crisis management in operating rooms (ORs) is crucial for patient safety. Despite their benefits, adherence to OR crisis checklists is often limited, highlighting the need for innovative solutions.

Objective: The objective of this study was to evaluate the efficacy of augmented reality (AR)-enhanced checklists in improving protocol adherence, compared to traditional paper checklists and no checklist scenarios during simulated OR crises.

Methods: This study was a randomized comparative efficacy study comparing the utility of AR checklists, paper checklists, and no checklist scenarios using 4 validated and simulated OR crises scenarios: asystolic cardiac arrest, air embolism, unexplained hypotension/hypoxia, and malignant hyperthermia. The study took place in a simulated OR setting and had applicability to the standard procedures in ORs, critical care units, and urgent care scenarios in the emergency department. To form the 24 OR teams, 50 professionals including 24 anesthesiologists, 24 nurses, 1 surgeon, and 1 scrub nurse from two academic hospitals were included. The primary outcome measured was the failure to adhere (FTA) rate for critical actions during simulated OR crises. Adherence was determined using retrospective video analysis involving 595 key processes evaluated across 24 surgical teams. Interrater reliability was assessed using a Cohen κ. Secondary outcomes included checklist usability and cognitive load, as measured by the low-frequency to high-frequency (LF/HF) ratio of the heart rate variability.

Results: The AR checklist group showed a significantly lower FTA rate (mean 15.1%, SD 5.77%) compared to the paper checklist (mean 8.32%, SD 5.65%; t_{23}=−2.08; P=.048) and the no checklist groups (mean 29.81%, SD 5.59%; t_{23}=−6.47; P<.001). The AR checklist also resulted in a higher LF/HF ratio for anesthesiologists ($F_{2,46}$=4.88; P=.02), showing a potential increase in the level of cognitive load. Survey data indicated positive receptions for both AR and paper checklists.

Conclusions: These results suggest that AR checklists could offer a viable method for enhancing adherence to critical care protocols. Although, further research is needed to fully assess their impact on clinical outcomes and to address any associated increase in cognitive load.

(JMIR XR Spatial Comput 2025;2:e60792) doi:10.2196/60792

KEYWORDS

augmented reality; operating room; crisis checklist; checklist; guideline adherence; quality improvement; patient safety; cardiac arrest; hypotension; hyperthermia; critical care; emergency department

Introduction

Unexpected crises in the operating room (OR), such as cardiac arrests or severe hemorrhages, create a critical situation in which surgical teams should deliver rapid and coordinated care with a time-sensitive order of actions listed in the OR crisis checklists [1-3]. Although these high-stakes, low-frequency crises may occur infrequently for any single practitioner, their cumulative incidence across hospitals underscores a significant challenge to patient safety and surgical outcomes [4-7]. The OR teams' ability to effectively manage these life-threatening complications depends on their preparedness in managing crises [8,9], training [10], and adherence to the validated crisis checklists [11].

Presurgical checklists are used before surgery to ensure correct patient identification and procedure planning. In contrast, crisis management checklists guide surgical teams during emergencies, helping them respond quickly to life-threatening situations. While both checklists improve safety, this study focuses specifically on crisis management checklists, which aim to support decision-making during critical events in the OR.

The lack of adherence to the checklists negatively impacts surgical mortality rates and overall hospital performance [12]. Evidence suggests that adherence to established best practices during these critical moments is varied and often associated with a decay in the retention of essential skills and knowledge over time [13-16]. In many instances, the use of surgical safety checklists was associated with a reduction in morbidity and mortality, and they were integrated as a new standard of care [17,18]. The dynamic and high-pressure nature of surgical emergencies requires not only adherence to protocols but also the ability to quickly access and use complex information under cognitively demanding conditions [19-21]. However, even though adherence to these checklists is crucial, the traditional paper ones are often difficult to use effectively in such intense scenarios [22-24]. The low adoption of checklists underscores the need for innovative approaches to using checklists that fit with surgical workflows, enhancing protocol adherence without disrupting the clinical focus.

Augmented reality (AR) technology, by relaying important procedural information directly into the clinicians' vision [25-28], can enhance protocol adherence in medical settings [29-33]. Initial applications of AR in medication management and emergency trauma care have shown promise in reducing errors and guiding clinicians through complex procedures with enhanced clarity and efficiency [34-38]. This evidence positions AR as a potential technology for improving adherence to medical protocols [39-41]. However, the effectiveness of and adherence to AR-enhanced surgical checklists during OR crises has not been thoroughly studied.

This study aims to evaluate the efficacy of AR-enhanced checklists in improving protocol adherence by surgical teams during simulated OR crises. By comparing outcomes with the traditional paper checklists and scenarios without a checklist, the research seeks to provide evidence on AR's utility to reduce the failure to adhere (FTA) rate for crucial procedural steps when managing surgical crises, ultimately improving patient outcomes in the OR. We hypothesize that the AR-enhanced checklists will significantly reduce the FTA rate for crucial procedural steps compared to traditional paper checklists and no checklist scenarios.

Methods

Study Design

This prospective within-subject study aimed to compare the impact of AR checklists, traditional paper checklists, and no checklist conditions on managing OR crises (Figure 1). A detailed outline of team participation and the methodological framework is included in Multimedia Appendix 1. The development and rationale behind the crisis checklists, guided by surgical safety standards, have been detailed in a previous publication [14]. Teams, including anesthesia staff, OR nurses, and a mock surgeon, faced simulated intraoperative crises with randomized scenario assignments and checklist types. Before the main investigation, a pilot study tested the scenario fidelity and the AR checklist's practicality. Paper checklists were provided in booklet form and placed near the anesthesia machine and the circulating nurse's station, mirroring their accessibility in actual ORs. A summary and the checklists are available in sections 1 - 3 of Multimedia Appendix 1.

Figure 1. Study overview diagram. (a) Checklists presented in an augmented reality interface using Microsoft HoloLens 2. (b) Study design scenarios including an augmented reality checklist, paper checklist, and no checklist.

Setups: The OR Checklists

We used OR crisis checklists for 4 critical scenarios: (1) asystolic cardiac arrest, (2) air embolism, (3) unexplained hypotension/hypoxia, and (4) malignant hyperthermia. These scenarios were derived from a comprehensive checklist development and testing process explained by Ziewacz et al [42] and were chosen for their clinical importance and feasibility for implementation in AR. Additionally, we followed the standardized approach used by Arriaga et al [14], which evaluated the efficacy of these checklists in improving adherence to lifesaving protocols through high-fidelity medical simulations. More details on the checklists and key processes evaluated to measure adherence to protocols can be found in section 3 of Multimedia Appendix 1.

Participants

Participants were recruited from 2 academic hospitals between October 2021, and September 2023. Each team comprised the anesthesia staff (including attending physicians and residents), OR nurses, one mock surgeon, and one scrub nurse, totaling 24 attending physicians and residents, 24 OR nurses, and one mock surgeon across 24 teams. Team formations were randomized. Each team dedicated an average of 3.5 hours within a single day to participate in a high-fidelity simulated OR environment. In the simulated OR, they encountered a series of crisis scenarios designed to test their adherence to critical and evidence-based practices. Recruitment of staff members was facilitated through sign-up sheets and random selection from those scheduled to work on designated study dates. Hospital departments arranged for staff to attend the simulation sessions instead of their regular workday. Hospital or department rules required that all anesthesia staff taking part had to have up-to-date certification

in advanced cardiac life support. Each participant only took part in one study session.

Ethical Considerations

Ethical approval for this study was obtained from the Ministry of Health, Kuwait (IRBl: SKU-219328). Informed consent was obtained from all participants prior to their involvement in the study. Participants were informed about the study's objectives, procedures, and their rights, including the ability to withdraw at any point without any repercussions. All data collected during the study were deidentified and stored securely to ensure participant confidentiality. Data were anonymized during analysis to protect privacy, and access was restricted to authorized personnel only. No monetary or nonmonetary compensation was provided to participants for their involvement in this study. Identifiable features of participants were not captured in any images or supplementary materials.

Primary Outcome: FTA rate

The primary outcome was the FTA rate for 47 key lifesaving processes outlined in Multimedia Appendix 1. Adherence was evaluated and scored as either yes or no by 2 physician reviewers from our team (AA and RG) who observed and scored recorded simulation sessions. These sessions were recorded as synchronized videos on 2 screens for a comprehensive review. To ensure the accuracy of adherence scoring, interrater reliability was assessed. Any disagreements or uncertainties in scoring were reviewed by third reviewers (CP, HS) and were resolved. The primary variables included the checklist group and the medical crisis scenario. The primary aspect of the study was the measured FTA rates.

Secondary Outcomes

Cognitive Load

We used a Polar chest strap to collect interbeat interval data from participants during scenarios with an accuracy of 1 millisecond. Previous studies have shown that a low-frequency to high-frequency (LF/HF) ratio extracted from heart rate variability is a validated proxy for cognitive load [43-45], particularly when collected using chest wraps [46]. We used NeuroKit2, a toolbox for neurophysiological signal processing [47], to extract the LF/HF ratio from data aggregated into a 1-minute time window.

Participant Satisfaction and Usability

To evaluate the ease of use and the perceived effectiveness of the AR and paper checklists, we administered a structured survey adopted from Arriaga et al [14]. The survey assessed participants' preparedness, ease of use, readability, willingness to use the checklist in real scenarios, and perceived impact on the clinical flow during emergencies. Responses were captured on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), providing insights into participants' attitudes and perceptions across various aspects of checklist usage.

Statistical Analysis

Participant characteristics were presented by descriptive statistical analysis, which reported the number and percentage of participants across different roles and years of experience. To assess the consistency in observational scoring, the agreement between two reviewers on the adherence scores was quantified using a Cohen κ. The Shapiro-Wilk test was used to evaluate the normality of the data distribution. ANOVA was used to compare the efficacy of interventions across 3 groups and post hoc analyses were conducted to examine the checklist's efficacy across various scenarios. Participant satisfaction and usability were analyzed using descriptive statistics and reporting means and SD. The statistical analyses were performed using SAS with all *P* values being 2-sided and a threshold for statistical significance set at $P<.05$.

Results

Participants

A total of 50 participants, forming 24 teams, took part in this study, which included anesthesiologists (n=14), anesthesia residents (n=10), OR nurses (n=24), a surgical resident (n=1), and a scrub nurse (n=1). All anesthesia residents were in the early stages of their careers with 0 - 2 years of experience, and OR nurses included a more diverse range of experience, spanning from 0 - 8 years. Each team contained 1 mock surgeon and 1 surgical assistant (scrub nurse), who attended as stand-in participants to the operative field without participating in decision-making or survey completion; these stand-in staff members were not counted as participants. Participants' years of experience are summarized in Table 1.

Table . Participant's role and their years of experience.

Role	Years of experience in specialty, n (%)			
	0 - 2	2 - 8	>8	Unknown
Anesthesiologist				
Attending physician (n=14)	0 (0)	7 (50)	7 (50)	0 (0)
Anesthesia resident (n=10)	10 (100)	0 (0)	0 (0)	0 (0)
Operating room nurse (n=24)	6 (25)	12 (50)	3 (12.5)	3 (12.5)
Surgical resident (n=1)	(1) 100	0 (0)	0 (0)	0 (0)
Scrub nurse (n=1)	0 (0)	1 (100)	0 (0)	0 (0)

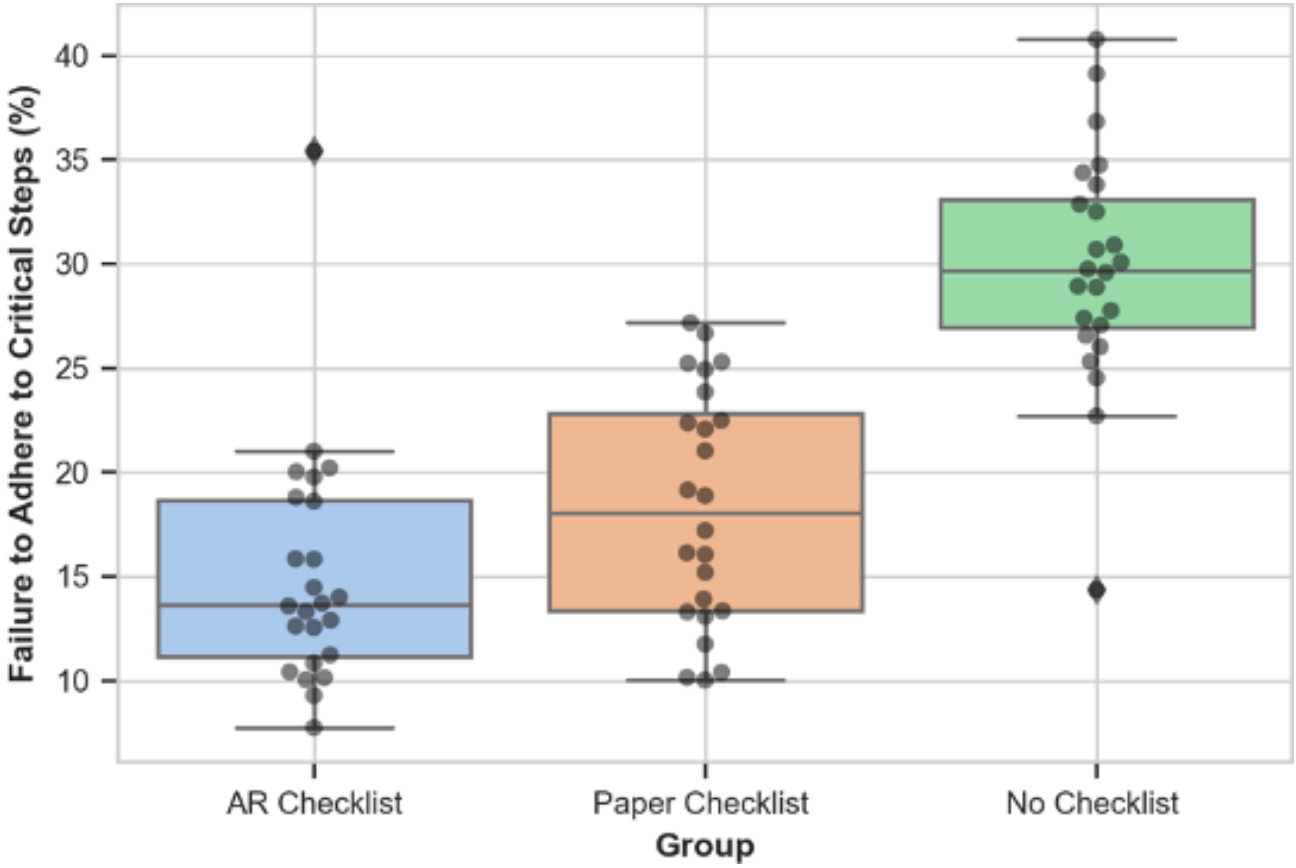
Adherence Rating

The assessment of adherence to key processes during the simulated scenarios demonstrated high interrater reliability among independent reviewer pairs, with Cohen κ values of ≥0.83 across all pairs. In instances where initial disagreement or uncertainty arose among the physician reviewers, consensus was reached through expert review with video replay. Out of a total of 595 key processes, evaluated across 24 teams for 25 key processes (excluding 8 key processes from one team that did not initiate the unexplained hypotension/hypoxia followed by an unstable bradycardia scenario), only 23 instances necessitated this expert review. The process of video replay facilitated immediate full agreement among all reviewers, highlighting the effectiveness of this approach in resolving ambiguities and ensuring accurate adherence assessment.

Comparing Groups Across All 4 Crisis Scenarios

ANOVA analysis showed significant differences in the FTA rate for critical steps among the 3 checklist groups ($F_{2,46}$=48.3; P<.001). Subsequent post hoc analysis showed the AR checklist group's mean FTA rate of 15.1% (SD 5.77%, 95% CI 13.50-16.70) was significantly lower than the paper checklist group's FTA rate of 18.32% (SD 5.65, 95% CI 16.75-19.89) and the no checklist group's FTA rate of 29.81% (SD 5.59, 95% CI 28.26-31.36). The AR group's FTA rate was significantly less than the no checklist group (t_{23}=−10.9; P<.001) and the paper checklist group (t_{23}=−2.08; P=.048). Moreover, the paper checklist group also had a significantly lower FTA rate compared to the no checklist group (t_{23}=−6.37; P<.001; Figure 2).

Figure 2. Failure to adhere to critical steps by condition type.



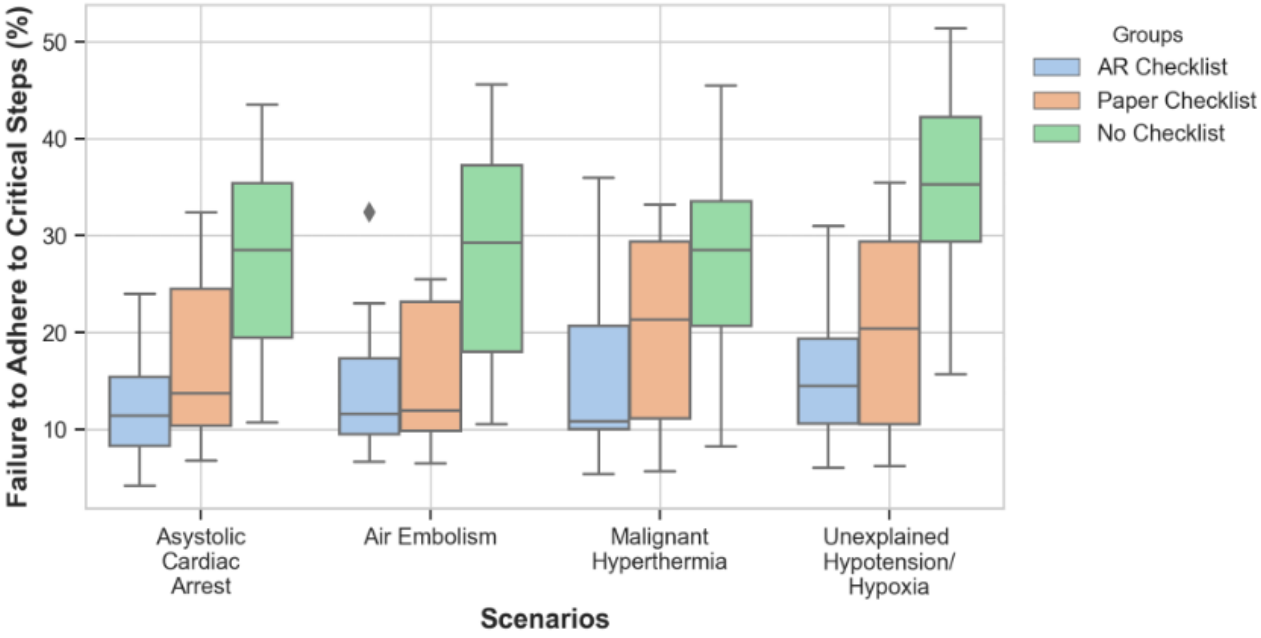
Comparing Groups for Individual Crisis Scenarios

Adherence to critical steps across various scenarios demonstrated significant differences among groups, with an ANOVA test showing distinct results for asystolic cardiac arrest ($F_{2,46}$=25.07; P<.001), air embolism ($F_{2,46}$=14.90; P<.001), malignant hyperthermia ($F_{2,46}$=12.33; P<.001), and unexplained hypotension/hypoxia ($F_{2,46}$=38.39; P<.001). Post hoc analyses indicated that, across these scenarios, the AR checklist group consistently exhibited significantly lower FTA rates compared to the no checklist group, with notable differences in asystolic cardiac arrest (t_{23}=−6.47; P<.001), air embolism (t_{23}=−4.45; P<.001), malignant hyperthermia (t_{23}=−4.79; P<.001), and unexplained hypotension/hypoxia (t_{23}=−10.57; P<.001). Comparisons between the AR and paper checklist groups were only significant for some scenarios, with slightly lower FTA rates for critical steps using the AR checklist in asystolic cardiac arrest (t_{23}=−2.65; P=.014) and unexplained hypotension/hypoxia (t_{23}=−2.10; P=.046). The paper checklist group also demonstrated significantly improved adherence over the no checklist condition in scenarios such as an air embolism (t_{23}=3.72; P<.001) and unexplained hypotension/hypoxia (t_{23}=5.40; P<.001; Figure 3).

While the AR checklist group demonstrated statistically significant differences in FTA rates compared to the paper

checklist group, it is important to note that this significance was observed by a narrow margin. Given the sample size, there remains the possibility that this effect could be influenced by chance, and further studies with larger sample sizes are necessary to confirm these findings.

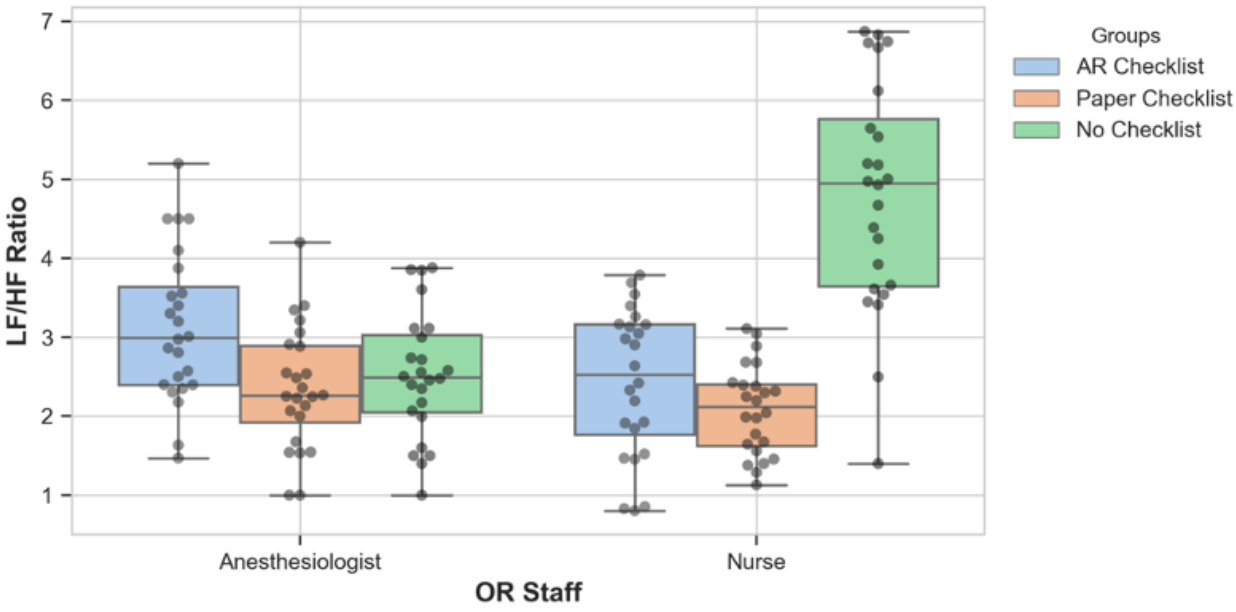
Figure 3. Failure to adhere to critical steps by scenario and group type. AR: augmented reality.



Cognitive Workload

For anesthesiologists, ANOVA results showed a significant effect of the checklist type on the LF/HF ratio ($F_{2,46}$=4.88; P=.02). In pairwise comparisons, the AR checklist group had a significantly higher LF/HF ratio compared to both the paper checklist and no checklist groups, suggesting a potential increase in cognitive load when using the AR checklist (P<.05; Figure 4). There was no significant difference in LF/HF ratio when comparing the paper checklist with no checklist groups, after adjusting for multiple comparisons. For nurses, the differences were significantly different ($F_{2,46}$=43.25; P<.001). The no checklist group had a significantly higher LF/HF ratio than the other two groups (P<.05). The AR checklist and paper checklist groups did not differ significantly.

Figure 4. Low-frequency to high-frequency ratio across operating room staff roles by checklist group. AR: augmented reality; LF/HF: low frequency to high frequency; OR: operating room.



Survey

Survey responses showed that both AR and paper checklist groups viewed their respective checklists positively (Table 2). Participants in the AR checklist group rated the checklist's ability to help them feel prepared during the emergency scenario at a mean Likert score of 4.5 (SD 0.75), and the paper checklist group rated this at 4.3 (SD 0.82), indicating no significant difference between the groups. Participants expressed a strong willingness to use the checklists in real-life situations, with the AR group scoring a 4.6 (SD 0.70) and the paper group scoring a 4.4 (SD 0.75). When considering the disruption to the clinical flow of the operative emergency, the AR checklist group reported less disruption with a mean score of 4.5 (SD 0.90) compared to the paper checklist group's score of 4.2 (SD 1.00).

Table . Questionnaire response data from participants on checklist usability.

Statement	AR[a] checklist group (n=48), mean (SD)	Paper checklist group (n=48), mean (SD)	*P* value
The checklist helped me feel better prepared during the emergency scenario.	4.5 (0.75)	4.3 (0.82)	.13
The checklist was easy to use.	4.4 (0.80)	4.2 (0.85)	.09
I would use this checklist if I were presented with this operative emergency in real life.	4.6 (0.70)	4.4 (0.75)	.03
The checklist did not disrupt the clinical flow of the operative emergency.	4.5 (0.90)	4.2 (1.00)	.04
If I were having an operation and experienced this intraoperative emergency, I would want the checklist to be used.	4.7 (0.55)	4.6 (0.60)	.18

[a]AR: augmented reality.

Discussion

Principal Findings

Our findings show that AR checklist groups had a superior adherence to critical steps in crises when compared to the paper checklist groups and groups who did not use any checklist. These findings highlight AR's potential to improve OR staff's adherence to predefined protocols and ultimately improve patient outcomes. This improvement suggests that sending critical and time-sensitive information to clinicians' and OR staff's field of view may help with faster and more precise decision-making in critical situations and emergencies. Considering a day-by-day improvement in technology, this will have the potential to set the ground for an extended and more effective AR checklist intervention in many other critical scenarios. This potential benefit is in line with a comparison of the AR checklist versus the traditional checklist in other health care applications [29,30]. The benefit of AR checklists, particularly in comparison with non-AR alternatives, underscores the technology's capacity to augment traditional safety measures.

It is also important to note that while the AR checklist group had a clear superiority over the no checklist group, the margin of improvement was modest when it was compared to the paper checklist group. In this comparison, the differences were not always statistically significant across different scenarios. These findings suggest that AR technology may not offer the same improvement in all clinical scenarios over the paper checklists. Considering the low sample size and extensive subgroup analysis, it is reasonable to suggest that AR's real-world application and its superiority over conventional methods warrant further examination. We also observed variation in team performance, as highlighted in Figure 1 of Multimedia Appendix 1. Some of this variation may be attributed to an order effect, where teams became more familiar with the simulation environment over time. This potential bias should be considered when interpreting the results, and future studies could include randomization or counterbalancing to mitigate this effect.

The feedback from participants indicated a high level of acceptance and perceived utility of AR checklists in crisis scenarios, pointing to the potential for AR to integrate effectively into surgical workflows. However, the nuanced performance improvements highlight the need for a tailored approach to technological integration in health care, where the specific context and user needs dictate the effectiveness of such alternatives [48-50]. The study's results align with broader trends in medical and high-risk industries, where checklists have long been recognized for their role in promoting adherence to best practices and enhancing outcomes [51-53]. Just as checklists have transformed safety protocols in aviation and nuclear power, AR checklists hold promise for surgical settings. Nonetheless, the adaptation of these tools in medicine, particularly in the high-stakes environment of the OR, requires careful consideration of design, implementation, and training to ensure they meet the unique demands of health care providers and patients.

A key consideration emerging from our research is the differential impact of AR on the cognitive load among OR staff. Anesthesiologists using the AR checklist have shown a higher LF/HF ratio, which may be associated with a higher level of cognitive load when compared to the paper and no checklist

groups. While we initially interpreted the higher LF/HF ratio in the AR checklist group as a sign of increased cognitive burden, it is also possible that this reflects heightened cognitive engagement. The AR checklist may stimulate more focused attention on the OR environment and monitoring, compared to the paper checklist, which could be perceived as more distracting. This alternative interpretation suggests that the AR condition may enhance attentional focus in a high-stakes environment, and further research is needed to clarify the relationship between LF/HF ratio and cognitive engagement.

It is an important finding that AR technology may improve adherence but simultaneously may add a cognitive burden [54,55] that adversely affects clinicians' behavior under cognitively demanding conditions. This variability in cognitive impact across different OR roles underscores the importance of designing AR applications that are tailored to the diverse needs and cognitive capacities of surgical teams. Future studies should also include qualitative methods to capture participants' experiences with AR and paper checklists. Combining this with quantitative data will provide a more complete understanding [56].

Limitations

This study has several limitations that should be considered. First, the study was conducted in a simulation setting that may not necessarily reflect the complexity of the OR environment. Second, our sample size was relatively small with a limited statistical power that prevented us from confidently performing subcategory analysis and extracting minor differences between groups. Larger studies with more diverse groups of clinicians and more scenario variability are needed to allow for subgroup analyses and to look for potential impacts on certain groups of clinicians or crisis scenarios. Third, the integration of AR technology into clinical practice raises questions about cost, accessibility, and the need for specialized training [57]. The development of best practices for the implementation and customization of AR checklists will be crucial to their successful adoption in surgical care. Last, we recognize that *P* values alone should not be taken as conclusive evidence of AR's superiority. The narrow statistical margin highlights the need for further validation through larger studies to confirm its efficacy.

Conclusion

Our study showed that the use of AR-enhanced checklists significantly improved adherence to critical procedural steps during simulated OR crises compared to both traditional paper checklists and scenarios without a checklist. These findings are promising as they may contribute to the patient's safety and outcomes. However, while the benefits of AR are promising, our findings also indicate a potential increase in cognitive load among clinicians, particularly anesthesiologists. Future studies should aim to optimize AR interfaces to minimize cognitive demands and validate these results in real-world settings. Addressing the balance between improved protocol adherence and cognitive load will be crucial for integrating AR effectively in high-stakes environments like the OR.

Conflicts of Interest

AG is the Medical Director of Ultrasight.

Multimedia Appendix 1
Supplementary materials on the development and application of augmented reality checklists for crisis management in clinical settings.
[DOCX File, 106 KB - xr_v2i1e60792_app1.docx]

References

1. Holzman RS, Cooper JB, Gaba DM, Philip JH, Small SD, Feinstein D. Anesthesia crisis resource management: real-life simulation training in operating room crises. J Clin Anesth 1995 Dec;7(8):675-687. [doi: 10.1016/0952-8180(95)00146-8] [Medline: 8747567]
2. Hepner DL, Arriaga AF, Cooper JB, et al. Operating room crisis checklists and emergency manuals. Anesthesiology 2017 Aug;127(2):384-392. [doi: 10.1097/ALN.0000000000001731] [Medline: 28604405]
3. Patel B, Thind A. Usability of mobile health apps for postoperative care: systematic review. JMIR Perioper Med 2020 Jul 20;3(2):e19099. [doi: 10.2196/19099] [Medline: 33393925]
4. Moorthy K, Munz Y, Forrest D, et al. Surgical crisis management skills training and assessment: a simulation[corrected]-based approach to enhancing operating room performance. Ann Surg 2006 Jul;244(1):139-147. [doi: 10.1097/01.sla.0000217618.30744.61] [Medline: 16794399]
5. Agarwala AV, Nurudeen S, Haynes AB. Perioperative checklists and handoffs. Adv Anesth 2015;33(1):191-217. [doi: 10.1016/j.aan.2015.07.011]
6. Harari RE, Dias RD, Kennedy-Metz LR, et al. Deep learning analysis of surgical video recordings to assess nontechnical skills. JAMA Netw Open 2024 Jul 1;7(7):e2422520. [doi: 10.1001/jamanetworkopen.2024.22520] [Medline: 39083274]
7. Dias RD, Harari RE, Zenati MA, et al. A clinician-centered explainable artificial intelligence framework for decision support in the operating theatre. Hamlyn Symp Med Robot 2024 Jun;16:35-36. [Medline: 39076411]
8. Harari R, Dias RD, Salas E, Unhelkar V, Chaspari T, Zenati M. Misalignment of cognitive processes within cardiac surgery teams. Hamlyn Symp Med Robot 2024 Jun;16:33-34. [Medline: 39081305]

9. Lin CC, Chen YP, Chiang CC, Chang MC, Lee OKS. Real-time streaming of surgery performance and intraoperative imaging data in the hybrid operating room: development and usability study. JMIR Med Inform 2020 Apr 22;8(4):e18094. [doi: 10.2196/18094] [Medline: 32209528]
10. Vogt L, Klasen M, Rossaint R, Goeretz U, Ebus P, Sopka S. Virtual reality tour to reduce perioperative anxiety in an operating setting before anesthesia: randomized clinical trial. J Med Internet Res 2021 Sep 1;23(9):e28018. [doi: 10.2196/28018] [Medline: 34252034]
11. Harari RE, Ahmadi N, Pourfalatoun S, Al-Taweel A, Shokoohi H. Clinician-AI collaboration for decision support in telemedicine: a randomized controlled trial study. 2024 Presented at: Proceedings of Conference on Cognitive and Computational Aspects of Situation Management 2023; Oct 16-19, 2023; Philadelphia, PA p. 81-89. [doi: 10.29007/9qxd]
12. Alfred MC, Cohen TN, Cohen KA, et al. Using flow disruptions to examine system safety in robotic-assisted surgery: protocol for a stepped wedge crossover design. JMIR Res Protoc 2021 Feb 9;10(2):e25284. [doi: 10.2196/25284] [Medline: 33560239]
13. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. N Engl J Med 2009 Jan 29;360(5):491-499. [doi: 10.1056/NEJMsa0810119] [Medline: 19144931]
14. Arriaga AF, Bader AM, Wong JM, et al. Simulation-based trial of surgical-crisis checklists. N Engl J Med 2013 Jan 17;368(3):246-253. [doi: 10.1056/NEJMsa1204720] [Medline: 23323901]
15. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med 2013 Oct 10;369(15):1434-1442. [doi: 10.1056/NEJMsa1300625] [Medline: 24106936]
16. Mcmullan RD, Urwin R, Gates P, Sunderland N, Westbrook JI. Are operating room distractions, interruptions and disruptions associated with performance and patient safety? A systematic review and meta-analysis. Int J Qual Health Care 2021 Apr 28;33(2):mzab068. [doi: 10.1093/intqhc/mzab068] [Medline: 33856028]
17. Babayan RK. Re: crisis checklists for the operating room: development and pilot testing. J Urol 2012 Feb;187(2):649. [doi: 10.1016/j.juro.2011.10.071] [Medline: 22237372]
18. Saxena S, Krombach JW, Nahrwold DA, Pirracchio R. Anaesthesia-specific checklists: a systematic review of impact. Anaesth Crit Care Pain Med 2020 Feb;39(1):65-73. [doi: 10.1016/j.accpm.2019.07.011] [Medline: 31374366]
19. Simmons WR, Huang J. Operating room emergency manuals improve patient safety: a systemic review. Cureus 2019 Jun 12;11(6):e4888. [doi: 10.7759/cureus.4888] [Medline: 31423368]
20. Gaba DM, Fish KJ, Howard SK, Burden A. Crisis Management in Anesthesiology: Elsevier Health Sciences; 2014:434.
21. Dryver E, Lundager Forberg J, Hård Af Segerstad C, Dupont WD, Bergenfelz A, Ekelund U. Medical crisis checklists in the emergency department: a simulation-based multi-institutional randomised controlled trial. BMJ Qual Saf 2021 Sep;30(9):697-705. [doi: 10.1136/bmjqs-2020-012740] [Medline: 33597283]
22. Weinger MB. Time out! Rethinking surgical safety: more than just a checklist. BMJ Qual Saf 2021 Aug;30(8):613-617. [doi: 10.1136/bmjqs-2020-012600] [Medline: 33758034]
23. Urban D, Burian BK, Patel K, et al. Surgical teams' attitudes about surgical safety and the surgical safety checklist at 10 years. Ann Surg Open 2021 Sep;2(3):e075. [doi: 10.1097/AS9.0000000000000075]
24. Neuhaus C, Spies A, Wilk H, Weigand MA, Lichtenstern C. "Attention everyone, time out!": safety attitudes and checklist practices in anesthesiology in Germany. A cross-sectional study. J Patient Saf 2021 Sep 1;17(6):467-471. [doi: 10.1097/PTS.0000000000000386] [Medline: 28574957]
25. Ebnali M, Fathi R, Lamb R, Pourfalatoun S, Motamedi S. Using augmented holographic UIs to communicate automation reliability in partially automated driving. 2020 Presented at: Automation Experience across Domains at CHI 2020; Apr 26-27, 2020 URL: https://ceur-ws.org/Vol-2700/paper5.pdf [accessed 2024-11-27]
26. Harari R, Al-Taweel A, Ahram T, Shokoohi H. Explainable AI and augmented reality in transesophageal echocardiography (TEE) imaging. Presented at: 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR); Jan 17-19, 2024; Los Angeles, CA p. 306-309. [doi: 10.1109/AIxVR59861.2024.00051]
27. Devoto L, Muscroft S, Chand M. Highly accurate, patient-specific, 3-dimensional mixed-reality model creation for surgical training and decision-making. JAMA Surg 2019 Oct 1;154(10):968-969. [doi: 10.1001/jamasurg.2019.2546] [Medline: 31433467]
28. Worlikar H, Coleman S, Kelly J, et al. Mixed reality platforms in telehealth delivery: scoping review. JMIR Biomed Eng 2023 Mar 24;8:e42709. [doi: 10.2196/42709] [Medline: 38875694]
29. Alismail A, Thomas J, Daher NS, et al. Augmented reality glasses improve adherence to evidence-based intubation practice. Adv Med Educ Pract 2019;10:279-286. [doi: 10.2147/AMEP.S201640] [Medline: 31191075]
30. Siebert JN, Ehrler F, Gervaix A, et al. Adherence to AHA guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: a randomized controlled trial. J Med Internet Res 2017 May 29;19(5):e183. [doi: 10.2196/jmir.7379] [Medline: 28554878]
31. Tsang KD, Ottow MK, van Heijst AFJ, Antonius TAJ. Electronic decision support in the delivery room using augmented reality to improve newborn life support guideline adherence: a randomized controlled pilot study. Simul Healthc 2022 Oct 1;17(5):293-298. [doi: 10.1097/SIH.0000000000000631] [Medline: 35102128]
32. Eves J, Sudarsanam A, Shalhoub J, Amiras D. Augmented reality in vascular and endovascular surgery: scoping review. JMIR Serious Games 2022 Sep 23;10(3):e34501. [doi: 10.2196/34501] [Medline: 36149736]

33. Puladi B, Ooms M, Bellgardt M, et al. Augmented reality-based surgery on the human cadaver using a new generation of optical head-mounted displays: development and feasibility study. JMIR Serious Games 2022 Apr 25;10(2):e34781. [doi: 10.2196/34781] [Medline: 35468090]
34. Kuyt K, Park SH, Chang TP, Jung T, MacKinnon R. The use of virtual reality and augmented reality to enhance cardio-pulmonary resuscitation: a scoping review. Adv Simul (Lond) 2021 Apr 12;6(1):11. [doi: 10.1186/s41077-021-00158-0] [Medline: 33845911]
35. Li C, Zheng B, Yu Q, Yang B, Liang C, Liu Y. Augmented reality and 3-dimensional printing technologies for guiding complex thoracoscopic surgery. Ann Thorac Surg 2021 Nov;112(5):1624-1631. [doi: 10.1016/j.athoracsur.2020.10.037] [Medline: 33275930]
36. Lawson J, Martin G, Guha P, et al. Effect of mixed reality on delivery of emergency medical care in a simulated environment: a pilot randomized crossover trial. JAMA Netw Open 2023 Aug 1;6(8):e2330338. [doi: 10.1001/jamanetworkopen.2023.30338] [Medline: 37639272]
37. Nickel F, Cizmic A, Chand M. Telestration and augmented reality in minimally invasive surgery: an invaluable tool in the age of COVID-19 for remote proctoring and telementoring. JAMA Surg 2022 Feb 1;157(2):169-170. [doi: 10.1001/jamasurg.2021.3604] [Medline: 34705030]
38. Wellens LM, Meulstee J, van de Ven CP, et al. Comparison of 3-dimensional and augmented reality kidney models with conventional imaging data in the preoperative assessment of children with Wilms tumors. JAMA Netw Open 2019 Apr 5;2(4):e192633. [doi: 10.1001/jamanetworkopen.2019.2633] [Medline: 31002326]
39. Ebnali M, Ahmadi N, Nabiyouni E, Karimi H. AI-powered human digital twins in virtual therapeutic sessions. Proc Int Symp Hum Factors Ergon Health Care 2023 Mar;12(1):1-4. [doi: 10.1177/2327857923121000]
40. Ebnali M, Paladugu P, Miccile C, et al. Extended reality applications for space health. Aerosp Med Hum Perform 2023 Mar 1;94(3):122-130. [doi: 10.3357/AMHP.6131.2023] [Medline: 36829279]
41. Burian BK, Ebnali M, Robertson JM, et al. Using extended reality (XR) for medical training and real-time clinical support during deep space missions. Appl Ergon 2023 Jan;106:103902. [doi: 10.1016/j.apergo.2022.103902] [Medline: 36162274]
42. Ziewacz JE, Arriaga AF, Bader AM, et al. Crisis checklists for the operating room: development and pilot testing. J Am Coll Surg 2011 Aug;213(2):212-217. [doi: 10.1016/j.jamcollsurg.2011.04.031] [Medline: 21658974]
43. Mizuno K, Tanaka M, Yamaguti K, Kajimoto O, Kuratsune H, Watanabe Y. Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity. Behav Brain Funct 2011 May 23;7:17. [doi: 10.1186/1744-9081-7-17] [Medline: 21605411]
44. Gabaude C, Baracat B, Jallais C, Bonniaud M, Fort A. Cognitive load measurement while driving. 2012 Presented at: Human Factors and Ergonomics Society Europe Chapter Annual Meeting; Oct 10-12, 2012; Toulouse, France p. 67-80 URL: https://hal.science/hal-01027475 [accessed 2024-12-30]
45. von Rosenberg W, Chanwimalueang T, Adjei T, Jaffer U, Goverdovsky V, Mandic DP. Resolving ambiguities in the LF/HF Ratio: LF-HF scatter plots for the categorization of mental and physical stress from HRV. Front Physiol 2017;8:360. [doi: 10.3389/fphys.2017.00360] [Medline: 28659811]
46. Villani V, Righi M, Sabattini L, Secchi C. Wearable devices for the assessment of cognitive effort for human–robot interaction. IEEE Sens J 2020 Nov 1;20(21):13047-13056. [doi: 10.1109/JSEN.2020.3001635]
47. Makowski D, Pham T, Lau ZJ, et al. NeuroKit2: a Python toolbox for neurophysiological signal processing. Behav Res Methods 2021 Aug;53(4):1689-1696. [doi: 10.3758/s13428-020-01516-y] [Medline: 33528817]
48. McCloskey K, Turlip R, Ahmad HS, Ghenbot YG, Chauhan D, Yoon JW. Virtual and augmented reality in spine surgery: a systematic review. World Neurosurg 2023 May;173:96-107. [doi: 10.1016/j.wneu.2023.02.068] [Medline: 36812986]
49. Rizzo MG Jr, Costello JP II, Luxenburg D, Cohen JL, Alberti N, Kaplan LD. Augmented reality for perioperative anxiety in patients undergoing surgery. JAMA Netw Open 2023 Aug 1;6(8):e2329310. [doi: 10.1001/jamanetworkopen.2023.29310] [Medline: 37589975]
50. Rosvall BR, Choby GW. Navigating the nose-the future of endoscopic endonasal surgical navigation. JAMA Otolaryngol Head Neck Surg 2024 Apr 1;150(4):326-327. [doi: 10.1001/jamaoto.2024.0025] [Medline: 38451546]
51. Ko HCH, Turner TJ, Finnigan MA. Systematic review of safety checklists for use by medical care teams in acute hospital settings--limited evidence of effectiveness. BMC Health Serv Res 2011 Sep 2;11. [doi: 10.1186/1472-6963-11-211] [Medline: 21884618]
52. Storesund A, Haugen AS, Flaatten H, et al. Clinical efficacy of combined surgical patient safety system and the World Health Organization's checklists in surgery: a nonrandomized clinical trial. JAMA Surg 2020 Jul 1;155(7):562-570. [doi: 10.1001/jamasurg.2020.0989] [Medline: 32401293]
53. Borchard A, Schwappach DLB, Barbir A, Bezzola P. A systematic review of the effectiveness, compliance, and critical factors for implementation of safety checklists in surgery. Ann Surg 2012 Dec;256(6):925-933. [doi: 10.1097/SLA.0b013e3182682f27] [Medline: 22968074]
54. Aitamurto T, Boin JB, Chen K, Cherif A, Shridhar S. The impact of augmented reality on art engagement: liking, impression of learning, and distraction. Presented at: Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry; Jul 15-20, 2018; Las Vegas, NV p. 153-171. [doi: 10.1007/978-3-319-91584-5_13]

55. Dixon BJ, Daly MJ, Chan H, Vescan AD, Witterick IJ, Irish JC. Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. Surg Endosc 2013 Feb;27(2):454-461. [doi: 10.1007/s00464-012-2457-3] [Medline: 22833264]
56. Weller JM, Merry AF, Robinson BJ, Warman GR, Janssen A. The impact of trained assistance on error rates in anaesthesia: a simulation-based randomised controlled trial. Anaesthesia 2009 Feb;64(2):126-130. [doi: 10.1111/j.1365-2044.2008.05743.x] [Medline: 19143687]
57. Borna A, Mousavi SZ, Fathollahzadeh F, Nazeri A, Harari RE. Applications of augmented and virtual reality in enhancing communication for individuals who are hard of hearing: a systematic review. Am J Audiol 2024 Oct 7:1-17. [doi: 10.1044/2024_AJA-24-00056] [Medline: 39374491]

Abbreviations

AR: augmented reality
FTA: failure to adhere
LF/HF: low-frequency to high-frequency
OR: operating room

Edited by L Riedemann, T Leung; submitted 21.05.24; peer-reviewed by B Robinson, K Bielka; revised version received 29.10.24; accepted 29.10.24; published 06.01.25.

Please cite as:
Ebnali Harari R, Altaweel A, Anderson E, Pozner C, Grossmann R, Goldsmith A, Shokoohi H
Augmented Reality in Enhancing Operating Room Crisis Checklist Adherence: Randomized Comparative Efficacy Study
JMIR XR Spatial Comput 2025;2:e60792
URL: https://xr.jmir.org/2025/1/e60792
doi:10.2196/60792

Integrating Extended Reality Into Primary Care Chronic Pain Programs via the REDOCVR Intervention: Real-World Implementation Feasibility and Usability Study

Jose Ferrer Costa[1,2,3,4], MD, MSc; Alexandra Fernandez Brusco[5]; Christian Torrecillas Camacho[5]; Elena Villabona Lopez[5]; Melanie Rodriguez Belloso[5]; Laura Villares Urgell[5]; Pablo Serrano Barrena[5]; Nuria Morán Bueno[5]

[1]Research and Innovation Department, Badalona Serveis Assistencials, Pl Pau Casals 1, Badalona, Spain

[2]Doctoral Programme in Health and Psychology, Universitat Oberta de Catalunya - UOC, Barcelona, Spain

[3]Research Support Unit, Metropolitana Nord, Jordi Gol i Gurina University Institute for Primary Health Care Research Foundation (IDIAPJGol), Mataró, Spain

[4]Multidisciplinary Research Group on Health and Society (GREMSAS) 2021 SGR 01484, Mataró, Spain

[5]Primary Care Services, Badalona Serveis Assistencials, Badalona, Spain

Corresponding Author:
Jose Ferrer Costa, MD, MSc
Research and Innovation Department, Badalona Serveis Assistencials, Pl Pau Casals 1, Badalona, Spain

Abstract

Background: Chronic pain management in public health services often struggles with limited engagement, emotional burden, and medication use. Extended reality (XR) shows promise in specialized settings, but evidence for codesign and integration into primary care remains limited.

Objective: This study aimed to examine the feasibility, usability, and real-world implementation of REDOCVR [RE (Reeducació), DOC (Dolor Crònic), VR (Virtual Reality)], an XR-supported psychoeducational program, and to explore preliminary clinical outcomes during its integration into chronic pain groups in public primary care centers.

Methods: This was a nonrandomized, hybrid type 2 phased implementation study conducted in 3 primary care centers in Catalonia, Spain. The intervention built on existing multidisciplinary psychoeducational chronic pain groups led by psychologists and physiotherapists. In collaboration with patients, XR modules were codesigned and incorporated to enhance mindfulness, cognitive reframing, and motor activation activities already established in routine care. In total, 8 weekly sessions included 15 - 20 minutes of this content, with a supervised medication tapering protocol included in later groups. Primary outcomes were implementation measures (adherence, tolerability, System Usability Scale, and satisfaction). Secondary outcomes included patient-reported clinical measures (Warwick-Edinburgh Mental Well-being Scale [WEMWBS], Hospital Anxiety and Depression Scale [HADS], Central Sensitization Inventory, and EuroQol – 5 Dimensions – 5 Levels) and medication changes, assessed at baseline, post-intervention, and 5-month follow-up.

Results: In total, 42 participants were enrolled, and 36 (85.7%) completed the intervention and all assessments. Adherence was high, and no serious adverse events occurred, with minimal cybersickness reported (5.6%). Patient usability was strong (mean 81.4, 95% CI 75.6 - 87.1), and overall satisfaction was high (mean 82.4, 95% CI 78.5 - 86.4). Professional usability was moderate (mean 59.1, 95% CI 51.6 - 66.5). Statistically significant improvements were observed in emotional well-being (Warwick-Edinburgh Mental Well-being Scale mean change 4.8, 95% CI 2.9 - 6.7; Cohen *d*=0.86), anxiety (HADS-A –2.5, 95% CI –3.8 to –1.2; Cohen *d*=0.66), and depression (HADS-D –1.6, 95% CI –2.5 to –0.7; Cohen *d*=0.62) (all *P*<.001). Mobility improved significantly (median change –1.0, 95% CI –1.0 to 0.0, *P*=.02), while Pain/Discomfort showed a nonsignificant trend (*P*=.08). Among tapering participants (n=22), mean use of benzodiazepines decreased by 71.7% and opioids by 41.8% at 5 months.

Conclusions: This study suggests that an XR-enhanced psychoeducational program can be incorporated into group-based chronic pain care within the public primary health care system. Exploratory improvements in emotional well-being, anxiety, depression, and reduced use of high-risk medications during supervised tapering indicate potential benefits, although causal inferences cannot be drawn given the feasibility design. These findings provide practical insights to inform refinement and progression to larger controlled studies evaluating scalability and long-term effects in routine primary care.

Trial Registration: ClinicalTrials.gov NCT06361706; http://clinicaltrials.gov/ct2/show/NCT06361706 and NCT06964360; http://clinicaltrials.gov/ct2/show/NCT06964360

(JMIR XR Spatial Comput 2025;2:e82858) doi:10.2196/82858

KEYWORDS

augmented reality; chronic pain; deprescriptions; feasibility studies; patient-centered care; patient education as topic; primary health care; virtual reality

Introduction

Chronic pain is common and disabling, with a recent systematic review estimating prevalence across European adults at roughly 21%, though individual studies range from 12% to 48% depending on criteria used [1]. In Spain, nationwide surveys report similar levels, with around 1 in 4 adults affected, and higher rates among women and in the 55 - 75 age group [2]. Beyond prevalence, chronic pain is consistently linked with functional impairment, lower quality of life, and high health care use [1]. At a societal level, productivity losses from sick leave and early retirement have been estimated to consume up to 4% of GDP in some countries [3].

Primary care is where clinicians first diagnose, treat, and follow most people with chronic pain over time [4,5], heavily relying on pharmacological treatment. In this setting, providers frequently prescribe opioids, nonsteroidal anti-inflammatory drugs (NSAIDs), antidepressants, anticonvulsants, and benzodiazepines despite limited long-term benefit and safety concerns, particularly with opioid-benzodiazepine combinations [6,7]. Multidisciplinary group programs that combine physical, psychological, and self-management strategies may provide a safer and more comprehensive approach with effectiveness supported by evidence [8]. Yet, their accessibility remains inconsistent, constrained by structural and organizational barriers [9].

Extended reality (XR), including virtual reality (VR) and augmented reality (AR), is being explored as an addition to nonpharmacological pain strategies. Recent systematic reviews and early trials in specialist centers suggest XR interventions may help reduce pain intensity, support emotional regulation, and improve engagement [10-12]. A recent randomized trial reported that VR-based therapy improved clinical outcomes as well as brain imaging markers in chronic back pain [13]. Most of this evidence, however, comes from tightly controlled contexts rather than everyday practice, underscoring the need for feasibility and implementation studies in primary care and community settings [14,15].

We developed REDOCVR [RE (Reeducació), DOC (Dolor Crònic), VR (Virtual Reality)] in 2023 within a public primary care network in Catalonia, Spain, where multidisciplinary group programs for chronic pain were already established and delivered by psychologists and physiotherapists in line with the regional pathway [16]. Instead of developing a separate intervention or testing XR in a laboratory, we included it into existing sessions, keeping the same structure, timing, and professional roles. Clinicians and patients codesigned content, bringing together XR experiences for guided mindfulness, psychoeducation, distraction, and interactive motor exercises.

This paper examines the feasibility and implementation of REDOCVR in public primary care. By doing so, it directly addresses recent calls for research on how XR interventions can be integrated beyond specialist centers and sustained in routine clinical practice [14,15]. Its innovation lies in embedding XR into established multidisciplinary group programs, codesigned with patients and clinicians, rather than creating a new parallel service. This pragmatic approach ensured the program was conceived and delivered entirely within real-life clinical settings, focusing on implementation outcomes relevant for scalability.

Methods

Study Design and Setting

We conducted a hybrid type 2, nonrandomized, phased implementation study conducted in public primary care centers (PCC). The rollout was structured in three phases; this paper reports on Phases 1 and 2, which examined feasibility and optimization, while Phase 3 is ongoing and will be reported separately.

Phases of Implementation

Phase 1: Codesign and Single-Site Feasibility

The first phase combined the participatory development process with a pilot group at PCC Apenins-Montigalà. Between September 2023 and February 2024, a core team of a psychologist, a physiotherapist, family physicians, a medical XR developer, technical partners, and patient representatives worked together to define content, delivery conditions, and safety procedures.

Professionals and patients cocreated the content to reflect the priorities of primary care, and immersive content was iteratively refined to ensure usability and acceptability. This included recording 360° videos with audio guides, reusing mindfulness modules from a previous project [17] (body scan and mindful breathing), and adapting the self-compassion module for the context of chronic pain. In parallel, AR hand-tracking exercises were designed in collaboration with professionals and patients to support safe, engaging movement, aligned with the routine interventions of the original program.

The pilot group (February-March 2024) completed eight 90-minute weekly sessions, with immersive content adjusted to the therapeutic goals of each session. Feasibility work at this stage focused on usability, acceptability, satisfaction, and iterative refinements of both content and technical procedures. XR was initially integrated into all weekly sessions in Phases 1 and 2; later adjustments are reported in the *Discussion* section.

Phase 2: Multisite Optimization

The program was expanded to include 2 additional PCCs while continuing at Apenins-Montigalà, comprising a total of 7 groups (3 at the original center, 2 at each new site). From the second group onward in each center, a supervised medication tapering protocol was introduced by family physicians, with scheduled reviews aligned to study timepoints.

Phase 3: Scale-Up

The program is currently extending to additional PCCs within Badalona Serveis Assistencials and to external institutions in the Catalan primary care system. This phase emphasizes readiness assessment, training of new professional teams, and preparation for replication beyond the initial network. The results from this phase will be reported separately.

Participants and Eligibility Criteria

Eligibility was limited to adults enrolled in chronic pain group programs routinely offered at the PCCs. Participants had experienced chronic pain for at least 3 months, most often related to fibromyalgia, with no restrictions on pain site or intensity. Exclusion criteria, based on diagnoses documented in the medical record, included severe psychiatric or cognitive disorders, epilepsy, major neurological or sensory impairments, contraindications for physical activity, or barriers that prevented regular attendance at group sessions.

Recruitment was coordinated by the physiotherapist and psychologist leading the groups at each center. Patients invited to join the chronic pain groups were informed that the program included the option of using XR as part of our research study. They could choose to participate with or without the XR component and were free to continue in the group even if they later decided to withdraw from the study. In routine practice, referrals to the chronic pain groups were generally based on patients presenting features such as central sensitization, emotional distress, kinesiophobia, or limited response to prior treatments, although these factors were not applied as formal inclusion criteria for the study. At a dedicated evaluation visit, eligibility was confirmed by the physiotherapist or psychologist responsible for the group, and written informed consent was obtained together with baseline data. All patients who were invited chose to participate.

As the intervention was embedded in standard care, no randomization or blinding was applied. No formal sample size calculation was conducted, as these phases focused on feasibility and to inform the design of subsequent groups. The sample size was pragmatic, determined by patient enrollment in chronic pain groups during the study period. A minimum of 30 participants was considered sufficient to capture feasibility issues and guide refinements for broader implementation.

Intervention

The intervention was delivered by psychologists, physiotherapists, and family physicians routinely leading group-based pain management programs. They received prior training and continuous technical support from the institutional innovation team, who had specific expertise in implementing XR in clinical care and codesigned the intervention content.

Immersive Dose and Delivery Conditions

XR content was applied for 15 - 20 minutes per session, with duration adjusted to the therapeutic focus and participants' tolerance. All modules were delivered under direct facilitator supervision. Headsets (Meta Quest 2 and 3) were preloaded with the required content, sanitized between uses, and configured individually. Quest 3 devices were prioritized for AR passthrough sessions when available, given their superior visual quality.

To support acclimatization among first-time users, XR elements were introduced in a stepwise sequence. Early sessions began with a passive, audio-guided VR body scan, after which participants progressed to psychoeducational and mindfulness modules, and later to AR passthrough motor exercises and interactive tasks. This gradual progression reflected strategies already tested in other XR initiatives within our institution, where it proved effective in improving tolerability and reducing cybersickness [18].

Final Intervention Components

The codesign process resulted in 2 integrated components that became the stable structure of REDOCVR sessions (see Figures 1 and 2). The XR content targeted 4 core domains relevant to chronic pain care in primary care: mindfulness and attentional focus; cognitive reframing and self-compassion; distraction for symptom regulation; and exercises to encourage safe, engaging movement.

Figure 1. Immersive content developed and implemented in REDOCVR Phase 1: Panels a–f illustrate the REDOCVR augmented reality (AR) application created for guided physiotherapy sessions, designed to encourage safe movement and reduce kinesiophobia. Tasks were implemented in passthrough mode with hand tracking: (a) basic target interaction, (b) activation of the user interface menu by facing the left palm toward the user, (c) manipulation of 3D objects to promote reaching and coordination, (d) mosquito interception game to train postural control and attentional focus, (e) spatial reasoning puzzle based on Tetris blocks, and (f) visual memory puzzle with image fragments. Panels g–i show Unity-based VR environments from the *Projecte Benestar* program, integrated into group psychology sessions to support nonpharmacological pain management: (g) body scan practice in a zen garden environment, (h) guided mindful breathing exercise at a virtual beach, and (i) psychoeducational module introducing the RAIN (Recognize, Allow, Investigate, Nurture) framework for emotion regulation.

Figure 2. Implementation of XR content in REDOCVR sessions. The left column illustrates psychology-led sessions using the 360° virtual reality (VR) videos to support mindfulness and emotional awareness. A supervising tablet (top) allowed facilitators to monitor the video being displayed in the headsets, while group sessions (middle and bottom) were conducted with live guidance. The right column illustrates physiotherapy-led sessions. The top panel shows VR-based physical activity using controllers. The middle panel displays the augmented reality (AR) interface of the REDOCVR application with a gamified motor task from the participant's perspective. The bottom panel depicts AR passthrough exercises using hand tracking during group movement practice. All sessions were supervised to ensure patient comfort and safety.

1. Psychology-focused sessions used VR modules to support psychoeducation, mindfulness, and emotional regulation. Depending on the session, participants worked with either the created 360° videos with audio guides ("Making peace with the pain" on natural beaches and a mindful dive in the Barcelona Aquarium) or Unity-based VR environments illustrating psychoeducational and cognitive reframing concepts. Modules were brief, noninteractive, and activated by simple hand-tracking. Together, they provided complementary approaches to emotional regulation, designed to be accessible to first-time XR users and flexible across sessions.
2. Physiotherapy-focused sessions integrated XR tools to encourage physical activation and body awareness. The REDOCVR AR app projected virtual objects into the physical space using passthrough and hand tracking, with tasks chosen by the physiotherapist according to functional goals: target interactions (touching floating objects), intercepting exercises requiring postural control (reacting to moving stimuli at 2 difficulty levels), and motor puzzles (shape-matching tasks). These modules incorporated elements of gamification, embodiment, and immersive presence to encourage motivation, although these factors were not directly assessed as outcomes. To provide a structured comparison, some sessions also used VR movement games from Immersive Oasis (a Spanish XR start-up). These games, operated with headset controllers rather than hand tracking, promoted functional movement while allowing evaluation of engagement and tolerability across different formats (AR vs VR, hand tracking vs controllers). Implementation fidelity was maintained by the standardized session scheme and calendar already established for the chronic pain groups, into which XR modules were integrated uniformly across sites.

All immersive content was available offline, preloaded in the headsets. Apps did not collect or transmit data; the only element

requiring Wi-Fi was the optional Reality Telling tablet app for synchronizing playback of 360° videos.

Outcome Measures

As this was an early-phase study, no explicit thresholds for feasibility or clinical outcomes were prespecified. The primary aim was to assess usability, tolerability, and feasibility of implementation under routine primary care conditions. The results were interpreted descriptively to guide adjustments in program delivery and inform the design of subsequent implementation phases, rather than to determine continuation or termination against predefined criteria.

Implementation Outcomes

Usability was assessed with 2 adaptations of the System Usability Scale (SUS), a validated instrument widely used in digital health research [19,20]. The patient version was reduced to 8 items, while the professional version retained the original 10 items with wording adapted to clinical use of XR. Both used 5-point Likert scales, with scores calculated following standard procedures and rescaled to a 0 - 100 range. This approach aligns with recommendations to adapt usability tools for real-world XR implementations [21] and is supported by evidence that removing alternating positive-negative phrasing improves clarity without compromising validity [22-24].

Satisfaction and engagement were evaluated with a 10-item questionnaire developed for this study, covering perceived usefulness, ease of use, content quality, emotional impact, and physical comfort, scored on a 5-point Likert scale with responses similarly rescaled to a 0 - 100 scale.

Adherence was defined as attendance at all scheduled sessions and completion of pre- or post-assessments. Most sessions were observed in person by the main investigator, with telephone debriefs where this was not possible. XR use and safety were monitored on a session-by-session basis, but without a structured or protocolized register; feedback was noted informally. Headset tolerability and safety were assessed with an 8-item questionnaire using 5-point Likert scales, covering comfort, fatigue, visual discomfort, dizziness, headache, and session interruption. Responses were analyzed descriptively at the item level rather than as a total score.

Qualitative feedback was obtained through a comments form, available at each session and again with the post-evaluation tests. The form contained a single free-text field without guiding questions. Responses were reviewed after each group to identify recurring perceptions. Informal remarks made during sessions were used to adjust delivery pragmatically but were not systematically examined.

Clinical Outcomes

Emotional well-being was measured with the 14-item Warwick-Edinburgh Mental Well-being Scale (WEMWBS), validated in Spanish samples with strong internal consistency (α=.90 - .93) [25,26].

Anxiety and depressive symptoms were assessed using the Hospital Anxiety and Depression Scale (HADS), Spanish version, which demonstrates strong internal consistency (α=.84 for anxiety, α=.85 for depression) and robust construct validity [27,28].

Central sensitization was evaluated with the Central Sensitization Inventory (CSI), scored 0 - 100, with scores ≥40 indicating clinically relevant sensitization. The Spanish version shows solid psychometric properties in clinical populations [29,30].

Health-related quality of life was measured with the 5-level EuroQol (EuroQol – 5 Dimensions – 5 Levels [EQ-5D-5L]), using Spanish population norms to derive utility indices [31,32].

Medication use was reviewed in scheduled visits with family physicians, who checked electronic prescriptions together with patients and confirmed actual doses and any additional drugs not recorded. This measure was applied only in groups that included the deprescribing component.

Data Management

Overview

We conducted assessments at baseline (month 0), post-intervention (month 2), and follow-up (month 5). Implementation outcomes were assessed at program completion, clinical outcomes at baseline and completion, and medication use at all 3 time points in groups where tapering was implemented. Questionnaires were administered on paper by the program facilitators, pseudonymized with unique alphanumeric codes, and transcribed into electronic spreadsheets. Data were stored on encrypted institutional servers with restricted access and regular backups, in compliance with Spanish data protection laws (Ley Orgánica de Protección de Datos y Garantía de Derechos Digitales 3/2018) and the European General Data Protection Regulation (EU 2016/679).

Statistical Analysis

Descriptive statistics were used to summarize sociodemographic data, clinical characteristics, and questionnaire responses. Continuous outcomes are presented as means with standard deviations or medians with interquartile ranges, according to distribution. Internal consistency was assessed for the adapted SUS questionnaires. Exploratory pre-post comparisons of clinical outcomes were conducted with paired-sample tests.

Reporting Guidelines

The study was reported in accordance with established guidelines. We followed the CONSORT extension for pilot and feasibility trials (see Checklist 1), the TIDieR checklist for intervention description (Checklist 2), RATE-XR for immersive technology research (Checklist 3), and iCHECK-DH for digital health implementation (Checklist 4).

Ethical Considerations

Ethical approval was obtained from the Institut Universitari d'Investigació en Atenció Primària Jordi Gol Research Ethics Committee for both study protocols (refs. 24/047-P and 24/211-ACps), which are registered at ClinicalTrials.gov (NCT06361706 and NCT06964360). The REDOCVR XR content was implemented only as nontherapeutic support within a group psychoeducational program. After review by the Spanish Agency of Medicines and Medical Devices, it was formally

classified as nonmedical device software, and no regulatory authorization was required.

All participants provided written informed consent before enrollment. Study data were deidentified prior to analysis, and no personal or health identifiers were collected by the XR software. Participants did not receive compensation for their participation. Figure 2 includes images where participants and clinicians are partially identifiable; written informed consent for publication of these images was obtained from all individuals depicted and archived by the research team.

Results

Sample Characteristics and Adherence

A total of 42 participants were enrolled across 7 intervention groups at 3 PCCs during 2024. Of these, 36 (85.7%) participants completed the full intervention and both pre- and post-assessments, with no missing outcome data among completers. Analyses of clinical outcomes were based on this final sample (n=36), with a mean age of 57.7 years (SD=9.36; range 40 - 79) and a predominance of women (97.2%, n=35). Adherence was high, with the 6 dropouts attributed to patients' logistical or scheduling conflicts, with no study-related reasons reported. Figure 3 illustrates a CONSORT-style participant flow diagram, while Table 1 details group composition across phases, from initial optimization to the later inclusion of supervised medication tapering.

Figure 3. Flow of participants through the REDOCVR intervention groups. CONSORT-style diagram illustrating enrollment, dropout, and analysis for the 7 nonrandomized groups implemented in 2024. A total of 42 participants were enrolled across 3 primary care centers, of whom 36 (85.7%) completed the full intervention and both pre- and post-assessments. Analyses of clinical outcomes were based on these 36 participants. Medication-related outcomes were analyzed separately in the subset of 22 participants who took part in groups that included supervised medication tapering.

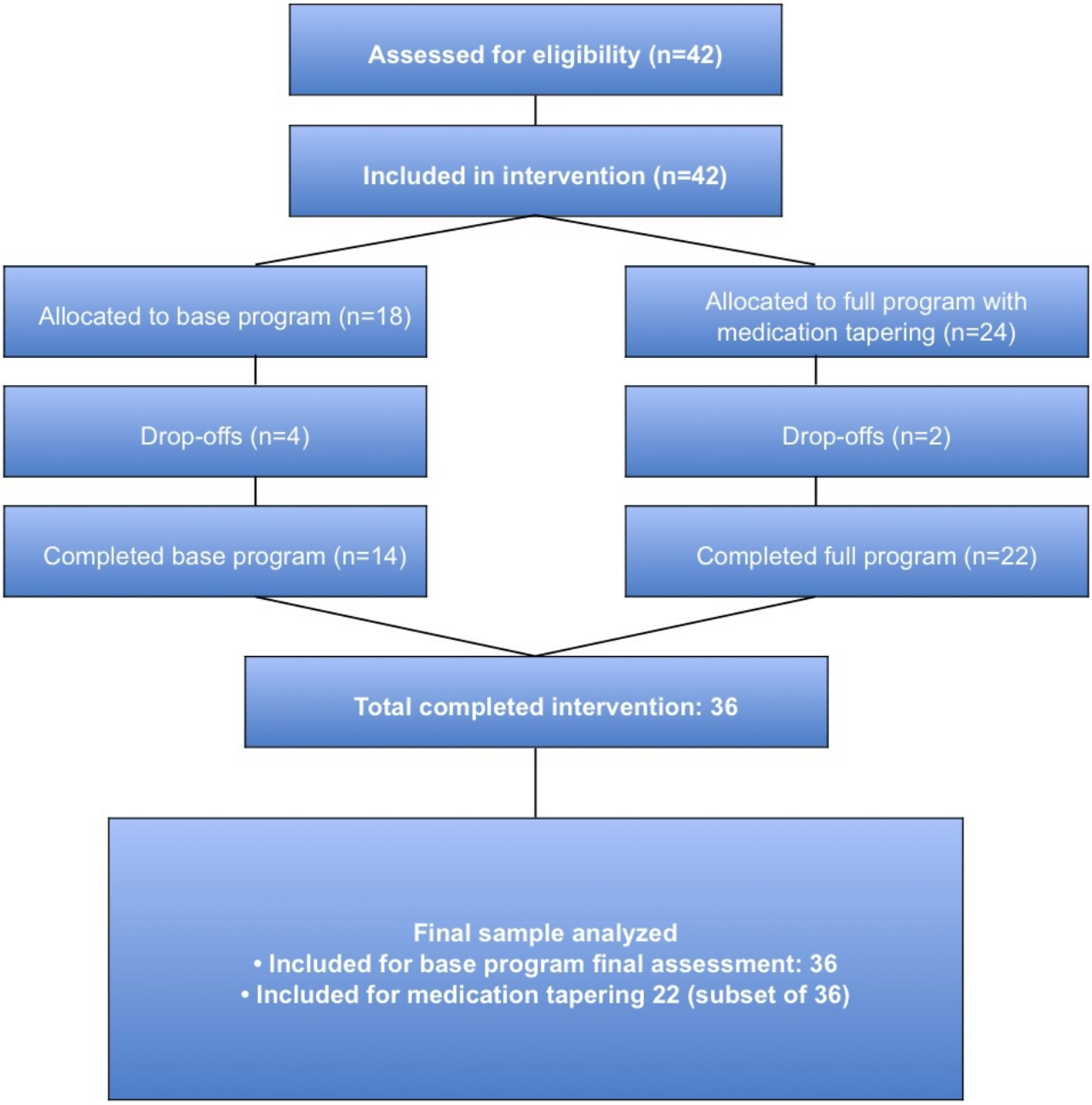


Table . Distribution of REDOCVR intervention groups and participants completed in 2024[a].

PCC[b]	Group	Phase	Implementation	Program sessions	Included (N=42), n	Completed post-evaluation (N=36), n	Medication tapering (N=22), n
Apenins-Montigalà	Group 1	Phase 1	Base program optimization	Feb-Mar 2024	6	4	-
Apenins-Montigalà	Group 2	Phase 2	Introduction of medication tapering	May-Jun 2024	6	6	6
Apenins-Montigalà	Group 3	Phase 2	Fully optimized program	Sept-Oct 2024	6	5	5
Progrés-Raval	Group 1	Phase 2	Base program only	Apr-May 2024	6	5	-
Progrés-Raval	Group 2	Phase 2	Fully optimized program	Sept-Nov 2024	6	6	6
Morera-Pomar	Group 1	Phase 2	Base program only	Apr-Jun 2024	6	5	-
Morera-Pomar	Group 2	Phase 2	Fully optimized program	Sept-Nov 2024	6	5	5

[a]Distribution of REDOCVR intervention groups completed in 2024 across 3 public primary care centers (PCCs) in Catalonia, Spain. The study followed a hybrid type 2, nonrandomized, phased implementation design. This paper reports on Phases 1 and 2, which focused on feasibility and optimization, while Phase 3 is ongoing and will be reported separately. The table shows participants' details by implementation phase, session completion, and inclusion of supervised medication tapering where applicable.

[b]PCC: primary care center.

Implementation Outcomes

System Usability

Patient-reported usability was high, with a mean SUS score of 81.4 (SD 17.1; 95% CI 75.6 - 87.1). Most participants (83%, n=30) scored above the standard benchmark of 68, indicating strong perceived usability. Item-level results, including English translations of the adapted questionnaire items, are provided in Table 2. Internal consistency of the adapted 8-item scale was acceptable (Cronbach α=.70).

SUS scores from 8 health care professionals (3 psychologists, 3 physiotherapists, 1 family physician, and 1 biomedical technician) averaged 59.06 (SD 8.96; 95% CI 51.6 - 66.5). Item-level ratings showed a mixed picture, with perceived contribution to patient care and clarity of simulation features scoring high, while system integration and need for support were rated lower. Internal consistency was low (Cronbach α=.28).

Table . Usability and satisfaction outcomes: patient and professional System Usability Scale (SUS), and patient satisfaction questionnaire.

Questions (English translation)	Mean (SD)	95% CI
(A) Patients—Adapted SUS (n=36, 8 items; Cronbach α=.70)[a]		
I think I would like to use this program frequently.	90.28 (22.58)	82.64 - 97.92
I think the program is unnecessarily complex.	75.69 (34.06)	64.17 - 87.22
I think the exercises are more engaging with the VR[b] headset than conventional methods.	82.64 (32.08)	71.78 - 93.49
I think the simulation options are clear and well integrated.	93.75 (19.25)	87.24 - 100.26
I think some options are difficult to follow.	75.69 (34.58)	63.99 - 87.4
I think people will learn to use this system easily.	81.25 (29.5)	71.27 - 91.23
I felt comfortable using this system.	89.58 (21.86)	82.19 - 96.98
I had to learn many things before being able to use the system.	61.81 (40.31)	48.17 - 75.44
Total patients SUS score	81.35 (17.1)	75.56 - 87.14
(B) Professionals—Adapted SUS (n=8, 10 items; Cronbach α=.28)[c].		
I believe this program will help improve the quality of patient care.	84.38 (12.94)	73.56 - 95.19
I believe the program interface is intuitive for the user.	75 (18.9)	59.2 - 90.8
I believe the program can be easily integrated into daily clinical practice.	53.13 (36.44)	22.66 - 83.59
I believe the simulation's features are clear and relevant to patient care.	81.25 (11.57)	71.57 - 90.93
Some features of this system may be difficult for patients to understand or use.	28.13 (20.86)	10.68 - 45.57
I believe patients will be able to learn to use this system easily.	68.75 (22.16)	50.22 - 87.28
I felt comfortable using this system for patient care.	78.13 (20.86)	60.68 - 95.57
You needed additional time to learn how to use this system correctly.	40.63 (32.56)	13.4 - 67.85
I believe the weight and design of the headset may be uncomfortable for some patients.	56.25 (34.72)	27.22 - 85.28
I believe using the headset and controllers could be difficult for some patients.	25 (18.9)	9.2 - 40.8
Total professionals SUS score	59.06 (8.96)	51.57 - 66.55
(C) Patients—Satisfaction questionnaire (ad hoc, n=36, 10 items; Cronbach α=.81)[d].		
Overall, how do you rate the VR experience during the chronic pain management sessions?	88.89 (13.94)	84.17 - 93.61
What is your opinion on the usefulness of VR for pain reduction?	82.64 (15.61)	77.36 - 87.92
Is the visual and image quality of the VR satisfactory for you?	88.19 (12.66)	83.91 - 92.48
Is interaction with the VR during sessions easy to understand and use?	87.5 (17.42)	81.6 - 93.4
How do you rate physical comfort while using VR?	80.56 (21.64)	73.23 - 87.88

Questions (English translation)	Mean (SD)	95% CI
Has VR improved your concentration and ability to manage pain?	73.61 (21.5)	66.34 - 80.89
What do you think about the variety of VR content available for your pain management sessions?	82.64 (18.73)	76.3 - 88.98
How do you rate the support and guidance you received during the VR sessions?	92.36 (14.42)	87.48 - 97.24
Is the duration of the VR sessions adequate for your goals?	75 (26.05)	66.19 - 83.81
Do you intend to continue using VR in your pain management sessions?	72.92 (25.62)	64.25 - 81.58
Totalpatients's satisfaction score	82.43 (11.69)	78.48 - 86.39

[a]Patient System Usability Scale (SUS, adapted 8-item version; n=36). Total and item-level mean scores are presented on a 0–100 scale.

[b]VR: virtual reality.

[c]Professional SUS (standard 10-item version; n=8 health care professionals: 3 psychologists, 3 physiotherapists, 1 family physician, and 1 biomedical technician). Total and item-level mean scores are presented on a 0–100 scale.

[d]Patient satisfaction questionnaire (ad hoc; n=36). Items were rated on a 0–100 scale and are shown as mean (SD). All questionnaire items were translated into English for reporting.

Satisfaction

Overall satisfaction was high, with a mean score of 82.4 (SD 11.7; 95% CI 78.4 - 86.4). The highest ratings were for support and guidance received (92.4), overall experience (88.9), and visual quality (88.2). Lower, though still positive, scores were reported for intention to continue use (72.9), session duration (75), and concentration and pain management (73.6). Item-level results with English translations are presented in Table 2.

Tolerability and Safety

Headset tolerability was generally high, with most participants disagreeing with statements indicating discomfort or adverse effects (Table 3). Dizziness (5.6%) and headache (5.6%) were rarely reported as frequent, and only 1 participant (2.8%) stopped a session due to discomfort. Fatigue in arms or hands was more common (33.3% agreement), likely reflecting the physical effort of interactive tasks rather than adverse effects. About half (50%) of participants rated the headset as comfortable for prolonged use. Overall, these findings suggest good tolerability with only occasional transient discomfort. Supervision confirmed correct use of XR without major incidents, and the open-ended feedback gathered during these sessions is described in the following section.

Table . Tolerability and safety of VR headsets (n=36)[a].

Question (English translation)	Strongly disagree (%)	Disagree (%)	Neutral (%)	Agree (%)	Strongly agree (%)
The headset was very heavy and uncomfortable	50	8.3	8.3	25	8.3
Using the headset and controls was very complicated	61.1	13.9	8.3	11.1	5.6
I felt fatigue in my arms and fingers	44.4	13.9	8.3	25	8.3
I felt visual discomfort	72.2	11.1	2.8	11.1	2.8
I felt dizzy	63.9	22.2	8.3	2.8	2.8
I experienced headache	77.8	8.3	8.3	2.8	2.8
I had to stop the simulation due to discomfort	83.3	8.3	5.6	0	2.8
I think it would be comfortable to use these glasses for a long time	8.3	13.9	27.8	22.2	27.8

[a]Tolerability and safety of virtual reality (VR) headsets during REDOCVR sessions among patients with chronic pain in public primary care (n=36). Responses are reported as the percentage of participants selecting each option on a 5-point Likert scale (strongly disagree to strongly agree). Questionnaire items were translated into English for reporting.

Qualitative and Comparative Feedback

Patient feedback, collected through the comments form (see Multimedia Appendix 1 for thematic analysis), underscored both therapeutic and practical aspects of the program. The 360° mindfulness videos and the self-compassion module were consistently described as the most meaningful. Many participants noted that the immersive format of body scan and breathing practices made these techniques easier to follow, while psychologists emphasized in their debriefings that these elements were particularly useful for consolidating skills introduced in other sessions. A few participants reported brief emotional discomfort during emotion-focused modules; these reactions were anticipated as part of the process, addressed by clinicians, and did not disrupt participation.

Patients also drew clear contrasts between modalities. AR exercises were described as intuitive and less tiring, with passthrough providing orientation that supported group flow. VR was seen as more immersive and distracting, though some participants reported temporary dizziness, fatigue, or visual strain after longer interactions. Across groups, patients consistently preferred hand-tracking over controllers, citing its natural interaction and shorter setup time (see Multimedia Appendix 2 for a video of the program sessions and content).

Professionals echoed some of these observations and noted additional logistical challenges. VR sessions required extra time to configure safety areas and adjust headsets individually. At least 1 physiotherapist described difficulty leading a group of 8 patients by himself, as sequential headset setup left participants waiting; these issues were less frequent when 2 or 3 professionals co-led with technical support. Setting up the connection for tablet-controlled synchronization of 360° videos was also seen as an added burden, further highlighting the need for streamlined logistics (see Multimedia Appendix 2 for a video of the program sessions and content).

Across groups, patients suggested extending the program, increasing session frequency, and considering access to VR content outside the clinical setting.

Clinical Outcomes

Normality assumptions determined the choice of test: paired-sample *t* tests were applied to normally distributed continuous outcomes (WEMWBS, HADS, CSI; Table 4), while Wilcoxon signed-rank tests were used for ordinal EQ-5D-5L domains (Table 5).

Table . Paired-sample *t* tests for normally distributed outcomes.

Variable	Mean difference (95% CI)	*P* value	Cohen *d*
Emotional wellbeing	4.83 (2.92 to 6.74)	<.001	0.86
HADS-Anxiety[a]	–2.47 (–3.75 to –1.20)	<.001	0.66
HADS-Depression	–1.64 (–2.53 to –0.75)	<.001	0.62
CSI[b]	–3.22 (–7.18 to 0.73)	0.1107	0.28

[a]HADS: Hospital Anxiety and Depression Scale.

[b]CSI: Central Sensitization Inventory.

Table . Wilcoxon signed-rank test results for EQ-5D-5L[a] dimensions

EuroQol (EQ-5D-5L) dimensions	HL[b] median change (95% CI)	*P* value	Effect size (r)	Interpretation
Mobility	–1.0 (–1.0 to 0.0)	0.0182	0.39	Significant improvement (moderate)
Self-care	0.0 (–0.5 to 0.0)	0.3374	0.16	No significant change (small)
Usual activities	0.0 (–0.5 to 0.5)	0.772	0.05	No change (very small)
Pain/Discomfort	–0.5 (–0.5 to 0.0)	0.084	0.29	Trend toward improvement (moderate)
Anxiety/Depression	0.0 (–0.5 to 0.0)	0.2869	0.18	No significant change (small)

[a]EQ-5D-5L: EuroQol 5-Dimension 5-Level questionnaire.

[b]HL: Hodges-Lehmann.

Table 6 presents pre-post values for clinical outcomes (n=36). Emotional well-being (WEMWBS) increased from 16.9 (SD 4.9) to 21.7 (SD 6.1), with a mean change of 4.8 (95% CI 2.9 - 6.7, *P*<.001, Cohen *d*=0.86). Anxiety symptoms (HADS-Anxiety) decreased from 13.1 (SD 4.3) to 10.7 (SD 4.3), mean change –2.5 (95% CI –3.8 to –1.2, *P*<.001, Cohen *d*=0.66), and depressive symptoms (HADS-Depression) from 10.6 (SD 3.4) to 9 (SD 4.2), mean change –1.6 (95% CI –2.5 to –0.7, *P*<.001, Cohen *d*=0.62). Central sensitization (CSI) declined slightly, from 60.7 (SD 14.2) to 57.5 (SD 13.4), mean change –3.2 (95% CI –7.2 to 0.7, *P*=.11, Cohen *d*=0.28).

Table . Clinical outcomes: descriptive statistics and pre–post differences (n=36)[a].

Variable	Mean (SD)	Median (IQR)	Range (Min-Max)	95% CI for mean
Descriptive statistics[b]				
Emotional wellbeing (Pre)	16.89 (4.92)	16.5 (6.25)	8 - 26	15.22 - 18.55
Emotional wellbeing (Post)	21.72 (6.14)	23.0 (8.25)	7 - 34	19.64 - 23.80
HADS-Anxiety (Pre)	13.14 (4.28)	13.5 (5.5)	4 - 20	11.69 - 14.59
HADS-Anxiety (Post)	10.67 (4.27)	10.5 (7.25)	4 - 20	9.22 - 12.11
HADS-Depression (Pre)	10.61 (3.43)	10.0 (5.0)	2 - 18	9.45 - 11.77
HADS-Depression (Post)	8.97 (4.23)	9.0 (5.0)	1 - 21	7.54 - 10.40
CSI[c] (Pre)	60.72 (14.18)	60.0 (18.0)	34 - 95	55.93 - 65.52
CSI (Post)	57.50 (13.42)	58.0 (15.25)	33 - 94	52.96 - 62.04
Mobility (Pre)	2.19 (0.92)	2.0 (1.25)	1 - 4	1.88 - 2.51
Mobility (Post)	1.78 (0.83)	2.0 (1.0)	1 - 4	1.50 - 2.06
Self-Care (Pre)	1.72 (0.66)	2.0 (1.0)	1 - 3	1.50 - 1.95
Self-Care (Post)	1.58 (0.81)	1.0 (1.0)	1 - 4	1.31 - 1.86
Usual activities (Pre)	2.47 (1.00)	2.5 (1.0)	1 - 5	2.13 - 2.81
Usual activities (Post)	2.42 (1.20)	3.0 (2.0)	1 - 5	2.01 - 2.82
Pain/Discomfort (Pre)	3.69 (0.75)	4.0 (1.0)	2 - 5	3.44 - 3.95
Pain/Discomfort (Post)	3.44 (0.73)	4.0 (1.0)	2 - 5	3.20 - 3.69
Anxiety/Depression (Pre)	3.31 (1.14)	3.0 (1.0)	1 - 5	2.92 - 3.69
Anxiety/Depression (Post)	3.06 (1.33)	3.0 (2.0)	1 - 5	2.61 - 3.51

[a]Clinical outcomes for patients with chronic pain participating in the REDOCVR program in public primary care (n=36).

[b]Descriptive statistics for all outcome measures, including mean (SD), median (IQR), range, and 95% CI.

[c]CSI, Central Sensitization Inventory.

For quality of life (EQ-5D-5L), Mobility improved significantly (median change –1.0, 95% CI –1.0 to 0.0, *P*=.02, *r*=0.39), while Pain/Discomfort showed a nonsignificant trend toward improvement (median change –0.5, 95% CI –0.5 to 0.0, *P*=.08, *r*=0.29). Self-Care, Usual Activities, and Anxiety/Depression domains showed no significant changes.

Medication Tapering

Among the 22 participants who followed the deprescription protocol, medication use declined across several categories (Table 7). Between baseline and the 5-month review, the largest mean reductions were in benzodiazepines (71.7%, 6 complete discontinuations), opioids (41.8%, 5 discontinuations), and muscle relaxants (48%). More modest decreases were seen for NSAIDs and non-opioid analgesics. Patterns for antidepressants and pain modulators were mixed, with some patients starting new treatments during follow-up.

Table . Medication tapering by therapeutic category at 2 and 5 months post-intervention[a].

Medication type	Timepoint (months)	Mean (%)	Median (%)	IQR (%)	Range (%)	N complete stops
NSAIDs[b]	2	29.86	33	45.75	0-100	1
NSAIDs	5	25.14	16.5	50	–100 to 100	2
Non-opioid analgesics	2	28.31	12.5	37.25	0-100	2
Non-opioid analgesics	5	30	12.5	49	0-100	2
Antidepressants	2	1.61	0	0	–100 to 100	3
Antidepressants	5	8.29	0	0	–100 to 100	6
Benzodiazepines	2	45	25	100	0-100	4
Benzodiazepines	5	71.7	100	45.75	0-100	6
Pain modulators	2	19.44	0	25	0-100	1
Pain modulators	5	34.22	25	50	0-100	2
Opioids	2	40.13	33	60	–100 to 100	5
Opioids	5	41.8	43	93	–100 to 100	5
Muscle relaxants	2	48	48	23	25-71	0
Muscle relaxants	5	48	48	23	25-71	0
Others	2	25	25	0	25-25	0
Others	5	100	100	0	100-100	1

[a]Medication tapering by therapeutic category at 2 and 5 months after participation in the REDOCVR program (n=22 patients with medication follow-up). Values represent percentage change in daily intake from baseline to each follow-up. Changes were calculated for each medication record and then aggregated by category. “N complete stops” indicates the number of participants who fully discontinued that medication. Negative values (eg, –100%) denote initiation of a medication not used at baseline. Data are reported as mean, median, interquartile range (IQR), and minimum-maximum percentage change for each category. “Others” refers to a single case of Versatis (lidocaine medicated plaster).

[b]NSAIDs: nonsteroidal anti-inflammatory drugs.

Discussion

Principal Findings

This study examined the early implementation of REDOCVR, an XR-enhanced group program for chronic pain in primary care. The findings suggest that immersive technologies can be integrated into existing care structures without disturbing routine delivery.

Satisfaction, tolerability, and usability indicators provide additional insight into how the program was received. Patients rated the sessions positively, reporting high satisfaction and minimal discomfort. Cybersickness was rare, and overall headset tolerance further supported the acceptability of XR in this population. The adapted 8-item SUS for patients reduced response burden while maintaining validity, consistent with prior analyses of SUS modifications [21-24,33]. In contrast, professional SUS scores showed low internal consistency, likely due to the small and heterogeneous sample. Such factors are known to reduce the stability of Cronbach α [34,35], whereas prior studies suggest that 10-item adaptations generally retain reliability [21,22]. Our findings, therefore, probably reflect contextual heterogeneity rather than flaws in the instrument itself.

An exploratory review of the open comments provided complementary insights into participants’ and professionals’ experiences with XR. Feedback suggests that XR was experienced less as a novel intervention in itself and more as a support that appeared to enhance engagement with the program’s therapeutic content. In psychology-led sessions, immersive tools supported mindfulness and self-compassion practices that some patients otherwise found difficult to understand. In physiotherapy-led sessions, AR passthrough helped participants perform movements in a safe and supervised way, while VR increased immersion and motivation, both encouraging patients to attempt physical activities they would normally avoid.

Exploratory clinical findings suggested improvements in emotional well-being, anxiety, and depression, with effect sizes in the moderate-to-large range. Statistically significant improvements were observed in Mobility on the EQ-5D-5L, while Pain/Discomfort showed only a nonsignificant positive trend. Among participants enrolled in the deprescription protocol, benzodiazepine, opioid, and NSAID intake decreased, and several participants achieved complete discontinuation. Although causality cannot be inferred, these results are consistent with the possibility that group-based interventions may complement structured deprescription strategies [36].

Comparison With Prior Work

While our findings on patient usability and clinical outcomes align with the existing literature [10,37], the primary contribution of this study is its focus on pragmatic implementation within a public primary care setting. We have encountered barriers in early adoption similar to other reports related to technical setup and workflow adjustments [38]. Prior reviews have noted that most VR interventions for pain concentrate on supporting specific aspects of health literacy or self-management rather than integration into routine care [39]. By embedding XR within an established group program rather than running it in parallel, REDOCVR addresses adoption challenges noted in recent reviews [14,15,40] and shows how immersive tools can fit into everyday service delivery. Delivering XR in group sessions brings clear advantages, including peer support, efficiency, and easier integration into existing care pathways, though it comes at the expense of individualized tailoring. Meta-analytic evidence in chronic pain indicates that group-based approaches can achieve outcomes comparable to individual formats [41]. This appears particularly relevant in the context of the public health care system, where resource optimization is not only desirable but often decisive for adoption.

The improvements in emotional well-being and functional outcomes align with the psychoeducational and behavioral mechanisms targeted, consistent with results from mindfulness- and acceptance-based approaches [8,13]. The reduction in medication use also mirrors outcomes from other multidisciplinary primary care interventions that combine patient education with supervised tapering [5]. What our findings add is preliminary evidence that such effects may be achievable when immersive tools are embedded into routine group care, rather than tested only in parallel or experimental conditions.

Limitations

The intervention was delivered to a modest number of groups within a supportive institutional setting, which may limit direct transferability to other contexts. Resources such as trained staff, headset availability, and organizational readiness facilitated implementation and should be considered when adapting the model elsewhere. Data collection on implementation fidelity was pragmatic. For instance, the dose of immersive exposure was guided by a time range (15 - 20 min) and adjusted for tolerance, but not strictly logged for each individual patient in every session. Adherence was assessed through attendance records, while adverse effects were monitored through informal therapist feedback and a single post-intervention questionnaire, rather than a structured, session-by-session register. The qualitative component was restricted to an exploratory review of written comments. While this offered useful contextual insights for program refinement, it was not intended as a formal qualitative study.

Other limitations are typical of early-phase implementation research, and exploratory clinical outcomes should be interpreted with caution. The absence of randomization and a control group limits causal inference. Generalizability is constrained by the modest and predominantly female sample (97%), which reflects the higher prevalence of chronic pain among women and the predominance of female participation in primary care group interventions in Spain [2,42]. This imbalance may restrict the applicability of findings to male patients. Participants also represented a heterogeneous chronic pain population, and we did not stratify by phenotype (nociceptive, neuropathic, or nociplastic), which may have masked differential responses. The group delivery model, while resource-efficient and well-suited to public primary care, inevitably restricts opportunities for individual tailoring [40]. The use of self-reported measures introduces potential recall and social desirability bias. Positive findings may have been partly influenced by a novelty effect. Finally, the 5-month follow-up is insufficient to establish the durability of changes over time.

Implications and Future Directions

Despite these constraints, the findings have already informed practical adjustments. Feedback from patients and professionals in Phases 1 and 2 reinforced the institutional support. Patients reported high satisfaction and usability, while professionals highlighted integration and workflow challenges. Combined with the clinical findings, these insights guided the Phase 3 refinements. The program was refined in collaboration with the full clinical team, leading to optimizations to improve feasibility.

In physiotherapy sessions, AR proved more suitable for larger groups, where supervision and interaction needed to be fluid, while VR was reserved for smaller groups. Both patients and professionals clearly favored hand-tracking, and all VR content was adapted accordingly, eliminating the need for controllers in the VR exercise games. Synchronizing 360° videos from the tablet required stable Wi-Fi, which was not always available. External routers were used to compensate, adding setup complexity. Efforts are underway to improve connectivity in PCCs and reduce facilitator workload.

Informed by early-phase findings and aligned with Catalan health system guidelines, the program will now follow the conventional format of 12 weekly sessions co-led by a psychologist and a physiotherapist, with selected sessions involving family physicians and nutritionists to reinforce healthy habits. XR use has been reorganized to fit within this structure, concentrated into 4 sessions—2 psychology-led (mindfulness and self-compassion) and 2 physiotherapy-led (motor activation). This targeted approach preserves therapeutic value while optimizing headset use and technical support across centers.

Future evaluation should focus on systematic data collection related to implementation fidelity. With XR use now organized into defined modules, session-by-session monitoring will allow more precise analysis of exposure, usability, and adverse effects across different modalities (eg, VR vs AR, passive vs interactive content). The expansion to additional centers and teams will be accompanied by structured pre-implementation focus groups and complementary usability strategies. Combined with established implementation frameworks, these steps are expected to generate the structured evidence needed to support broader adoption. By linking fidelity monitoring with pragmatic implementation, REDOCVR may help close the gap between immersive research and everyday chronic pain care.

Conclusions

The first 2 phases of REDOCVR suggest that a person-centered, codesigned XR program can be incorporated in primary care chronic pain groups under real-world conditions. The intervention showed high adherence and favorable usability, and both patients and professionals valued its contribution to engagement and session delivery. Exploratory improvements in emotional well-being and reductions in high-risk medication use with supervised tapering are promising but should be interpreted with caution given the feasibility design.

This study is, to our knowledge, the first to develop and evaluate an XR program entirely within primary public health care. The findings offer practical guidance for refining XR-based interventions and preparing for wider scale-up. Ongoing work on Phase 3 across additional centers will further clarify long-term outcomes, scalability, and adaptability to different settings.

Acknowledgments

We thank the patients who participated in REDOCVR, the primary care teams at Badalona Serveis Assistencials, and the collaborators Reality Telling and Immersive Oasis. ChatGPT (GPT-5, OpenAI, 2025) was used exclusively for language refinement. All outputs were critically reviewed, edited, and approved by the authors to ensure accuracy and fidelity to the study data. This work was partially funded by the Spanish Ministry of Health under the Primary and Community Care Strategy, financed in 2023 and 2024 following agreement of the Interterritorial Council of the Spanish National Health System (budget line 26.07.313B.456). The project is registered as P07.1: Reinventing the management of chronic pain with virtual reality, focusing on coordinated interventions for fibromyalgia and chronic pain in primary care. Additional costs not covered by the Ministry, including dissemination and publication expenses, were supported by Badalona Serveis Assistencials. The funders had no involvement in the study design, data collection, analysis, interpretation of results, or the writing of the manuscript.

Data Availability

The anonymized datasets generated and analyzed during this study are available from the corresponding author upon reasonable request and subject to approval by the Ethics Committee.

Authors' Contributions

Conceptualization: AFB, CTC, JFC, NMB
Data curation: JFC
Formal analysis: JFC
Investigation: AFB, CTC, EVL, JFC, LVU, MRB, PSB
Methodology and validation: AFB, CTC, JFC, NMB
Project administration: JFC
Resources: JFC
Software and VR content development: JFC
Writing–original draft preparation: JFC
Writing–review and editing: AFB, CTC, EVL, JFC, LVU, NMB

Conflicts of Interest

None declared.

Multimedia Appendix 1
Thematic analysis of patient feedback with illustrative quotes.
[PDF File, 77 KB - xr_v2i1e82858_app1.pdf]

Multimedia Appendix 2
REDOCVR BSA overview 2025: project content sample and real-life use in primary care.
[MP4 File, 123500 KB - xr_v2i1e82858_app2.mp4]

Checklist 1
CONSORT Checklist (Feasibility Extension).
[PDF File, 64 KB - xr_v2i1e82858_app3.pdf]

Checklist 2
TIDieR Checklist.
[PDF File, 61 KB - xr_v2i1e82858_app4.pdf]

Checklist 3
RATE-XR Checklist.
[PDF File, 66 KB - xr_v2i1e82858_app5.pdf]

Checklist 4
iCHECK-DH Checklist.
[PDF File, 63 KB - xr_v2i1e82858_app6.pdf]

References

1. Rometsch C, Martin A, Junne F, Cosci F. Chronic pain in European adult populations: a systematic review of prevalence and associated clinical features. Pain 2025 Apr 1;166(4):719-731. [doi: 10.1097/j.pain.0000000000003406] [Medline: 40101218]
2. Dueñas M, De Sola H, Salazar A, Esquivia A, Rubio S, Failde I. Prevalence and epidemiological characteristics of chronic pain in the Spanish population. Results from the pain barometer. Eur J Pain 2025 Jan;29(1):e4705. [doi: 10.1002/ejp.4705] [Medline: 39046161]
3. Stubhaug A, Hansen JL, Hallberg S, Gustavsson A, Eggen AE, Nielsen CS. The costs of chronic pain—long-term estimates. Eur J Pain 2024 Jul;28(6):960-977. [doi: 10.1002/ejp.2234] [Medline: 38214661]
4. Sociedad española del dolor. El dolor crónico en España. 2022. URL: https://www.sen.es/saladeprensa/pdf/Link387.pdf [accessed 2025-08-11]
5. Bults RM, van Dongen JM, Ostelo R, Nijs J, Keizer D, van Wilgen CP. Effectiveness of a primary care multidisciplinary treatment for patients with chronic pain compared with treatment as usual. J Clin Med 2023 Jan 22;12(3):885. [doi: 10.3390/jcm12030885] [Medline: 36769533]
6. Russo M, Nevitt S, Santarelli D, Eldabe S, Duarte RV. Systematic review and meta-analysis of conventional medical management in a patient population with refractory chronic pain suitable to receive a spinal cord stimulation system. Pain Med 2025 Jun 1;26(6):337-347. [doi: 10.1093/pm/pnaf004] [Medline: 39898812]
7. Busse JW, Wang L, Kamaleldin M, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. JAMA 2018 Dec 18;320(23):2448-2460. [doi: 10.1001/jama.2018.18472] [Medline: 30561481]
8. Robles E, Blanco I, Díez G, Vázquez C. Mindfulness-based stress reduction for chronic pain: enhancing psychological well-being without altering attentional biases towards pain faces. Eur J Pain 2025 Feb;29(2):e4714. [doi: 10.1002/ejp.4714] [Medline: 39180265]
9. Ashcraft LE, Hamm ME, Omowale SS, et al. The perpetual evidence-practice gap: addressing ongoing barriers to chronic pain management in primary care in three steps. Front Pain Res (Lausanne) 2024;5:1376462. [doi: 10.3389/fpain.2024.1376462] [Medline: 39439739]
10. Brown L, DiCenso-Fleming T, Ensign T, Boyd AJ, Monaghan G, Binder DS. Chronic pain education delivered with a virtual reality headset in outpatient physical therapy clinics: a multi-site exploratory trial. Am J Transl Res 2023;15(5):3500-3510. [Medline: 37303618]
11. Kumar V, Vatkar AJ, Kataria M, Dhatt SS, Baburaj V. Virtual reality is effective in the management of chronic low back ache in adults: a systematic review and meta-analysis of randomized controlled trials. Eur Spine J 2024 Feb;33(2):474-480. [doi: 10.1007/s00586-023-08040-5] [Medline: 37979068]
12. Sarkar TD, Edwards RR, Baker N. The feasibility and effectiveness of virtual reality meditation on reducing chronic pain for older adults with knee osteoarthritis. Pain Pract 2022 Sep;22(7):631-641. [doi: 10.1111/papr.13144] [Medline: 35750655]
13. Čeko M, Baeuerle T, Webster L, Wager TD, Lumley MA. The effects of virtual reality neuroscience-based therapy on clinical and neuroimaging outcomes in patients with chronic back pain: a randomized clinical trial. Pain 2024 Aug 1;165(8):1860-1874. [doi: 10.1097/j.pain.0000000000003198] [Medline: 38466872]
14. Ding ME, Traiba H, Perez HR. Virtual reality interventions and chronic pain: scoping review. J Med Internet Res 2025 Feb 18;27:e59922. [doi: 10.2196/59922] [Medline: 39965778]
15. Hess CW, Logan DE, Rosenbloom BN, et al. Developing a core outcome set for pediatric and adult acute and chronic pain extended reality trials: Delphi consensus-building process. J Med Internet Res 2025 May 23;27:e58947. [doi: 10.2196/58947] [Medline: 40408138]
16. Generalitat de catalunya. Fibromiàlgia – Canal Salut Generalitat de Catalunya. 2025. URL: https://canalsalut.gencat.cat/ca/salut-a-z/f/fibromialgia/ [accessed 2025-08-11]
17. Ferrer Costa J, Moran N, Garcia Marti C, Colmenares Hernandez LJ, Radu Ciorba Ciorba F, Ciudad MJ. Immediate impact of an 8-week virtual reality educational program on burnout and work engagement among health care professionals: pre-post pilot study. JMIR XR Spatial Comput 2024;1:e55678. [doi: 10.2196/55678]
18. Ferrer Costa J, Ciudad MJ, González Abad M, Rodríguez García JL. Virtual reality in cognitive rehabilitation: a pilot study on usability and preliminary clinical outcomes in seniors with mild cognitive impairment. J Med Ext Reality 2025;2(1):30-42. [doi: 10.1089/jmedxr.2024.0051]

19. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis:189-194 URL: https://www.taylorfrancis.com/books/9781498710411 [doi: 10.1201/9781498710411-35]
20. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
21. Zanatta F, Giardini A, Pierobon A, D'Addario M, Steca P. A systematic review on the usability of robotic and virtual reality devices in neuromotor rehabilitation: patients' and healthcare professionals' perspective. BMC Health Serv Res 2022 Apr 20;22(1):523. [doi: 10.1186/s12913-022-07821-w] [Medline: 35443710]
22. Sauro J, Lewis JR. When designing usability questionnaires, does it hurt to be positive? 2011 May 7 Presented at: CHI '11; May 7, 2011; Vancouver, BC, Canada p. 2215-2224 URL: https://dl.acm.org/doi/proceedings/10.1145/1978942 [doi: 10.1145/1978942.1979266]
23. Kortum PT, Bangor A. Usability ratings for everyday products measured with the system usability scale. Int J Hum Comput Interact 2013 Jan;29(2):67-76. [doi: 10.1080/10447318.2012.681221]
24. Lewis JR, Sauro J. Revisiting the factor structure of the system usability scale. J Usability Stud 2017;12(4):183-192 [FREE Full text]
25. Castellví P, Forero CG, Codony M, Vilagut G, Brugha TS, Ferrer M, et al. The Spanish version of the Warwick–Edinburgh mental well-being scale: validation and psychometric properties in a representative sample. Qual Life Res 2014;23(4):1279-1284. [doi: 10.1007/s11136-013-0513-7]
26. López MA, Gabilondo A, Codony M, et al. Adaptation into Spanish of the Warwick-Edinburgh mental well-being scale (WEMWBS) and preliminary validation in a student sample. Qual Life Res 2013 Jun;22(5):1099-1104. [doi: 10.1007/s11136-012-0238-z] [Medline: 22836376]
27. Herrero MJ, Blanch J, Peri JM, De Pablo J, Pintor L, Bulbena A. A validation study of the hospital anxiety and depression scale (HADS) in a Spanish population. Gen Hosp Psychiatry 2003;25(4):277-283. [doi: 10.1016/s0163-8343(03)00043-4] [Medline: 12850660]
28. Quintana JM, Padierna A, Esteban C, Arostegui I, Bilbao A, Ruiz I. Evaluation of the psychometric characteristics of the Spanish version of the hospital anxiety and depression scale. Acta Psychiatr Scand 2003 Mar;107(3):216-221. [doi: 10.1034/j.1600-0447.2003.00062.x] [Medline: 12580829]
29. Mayer TG, Neblett R, Cohen H, et al. The development and psychometric validation of the central sensitization inventory. Pain Pract 2012 Apr;12(4):276-285. [doi: 10.1111/j.1533-2500.2011.00493.x] [Medline: 21951710]
30. Cuesta-Vargas AI, Roldan-Jimenez C, Neblett R, Gatchel RJ. Cross-cultural adaptation and validity of the Spanish central sensitization inventory. SpringerPlus 2016;5(1):1837. [doi: 10.1186/s40064-016-3515-4] [Medline: 27818875]
31. Hernandez G, Garin O, Pardo Y, et al. Validity of the EQ-5D-5L and reference norms for the Spanish population. Qual Life Res 2018 Sep;27(9):2337-2348. [doi: 10.1007/s11136-018-1877-5] [Medline: 29767329]
32. Martí-Pastor M, Pont A, Ávila M, et al. Head-to-head comparison between the EQ-5D-5L and the EQ-5D-3L in general population health surveys. Popul Health Metr 2018 Aug 16;16(1):14. [doi: 10.1186/s12963-018-0170-8] [Medline: 30115092]
33. Lewis JR, Sauro J. Can I leave this one out? The effect of dropping an item from the SUS. J Usability Stud 2017;13(1):38-41. [doi: 10.5555/3173069.3173073]
34. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55. [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
35. Trizano-Hermosilla I, Alvarado JM. Best alternatives to Cronbach's alpha reliability in realistic conditions: congeneric and asymmetrical measurements. Front Psychol 2016;7:769. [doi: 10.3389/fpsyg.2016.00769] [Medline: 27303333]
36. Wang Y, Wilson DL, Fernandes D, et al. Deprescribing strategies for opioids and benzodiazepines with emphasis on concurrent use: a scoping review. JCM 2023;12(5):1788. [doi: 10.3390/jcm12051788]
37. Tuck N, Pollard C, Good C, et al. Active virtual reality for chronic primary pain: mixed methods randomized pilot study. JMIR Form Res 2022 Jul 13;6(7):e38366. [doi: 10.2196/38366] [Medline: 35830224]
38. Porta X, Nieto R, Serrat M, Bourdin-Kreitz P. Exploring user perspectives on virtual reality as a therapeutic tool for chronic pain management. Musculoskeletal Care 2025 Jun;23(2):e70101. [doi: 10.1002/msc.70101] [Medline: 40214985]
39. Skidmore N, Ryan C, Mankelow J, Bradford C, Graham A, Martin D. Exploring the potential of virtual reality for the self-management of chronic pain: a scoping review of its use to address health literacy. Musculoskelet Sci Pract 2024 Aug;72:102962. [doi: 10.1016/j.msksp.2024.102962] [Medline: 38703701]
40. Slatman S, Staal JB, van Goor H, Ostelo R, Soer R, Knoop J. Limited use of virtual reality in primary care physiotherapy for patients with chronic pain. BMC Musculoskelet Disord 2024 Feb 22;25(1):168. [doi: 10.1186/s12891-024-07285-5] [Medline: 38388377]
41. Niknejad B, Bolier R, Henderson CR Jr, et al. Association between psychological interventions and chronic pain outcomes in older adults: a systematic review and meta-analysis. JAMA Intern Med 2018 Jun 1;178(6):830-839. [doi: 10.1001/jamainternmed.2018.0756] [Medline: 29801109]
42. Cantero-Braojos M, Cabrera-León A, López-González MA, Saúl LA. Intervención grupal desde un enfoque sensoriomotriz para reducir la intensidad del dolor crónico. Atención Primaria 2019 Mar;51(3):162-171. [doi: 10.1016/j.aprim.2017.07.006]

Abbreviations

AR: augmented reality
CSI: Central Sensitization Inventory
EQ-5D-5L: EuroQol – 5 Dimensions – 5 Levels
HADS: Hospital Anxiety and Depression Scale
NSAIDs: nonsteroidal anti-inflammatory drugs
PCC: primary care center
REDOCVR: RE (Reeducació), DOC (Dolor Crònic), VR (Virtual Reality)
SUS: System Usability Scale
VR: virtual reality
WEMWBS: Warwick-Edinburgh Mental Well-being Scale
XR: extended reality

Edited by T Leung; submitted 22.08.25; peer-reviewed by D Ife-Tugbiyele, JG Armas; revised version received 01.10.25; accepted 13.10.25; published 31.10.25.

Please cite as:
Ferrer Costa J, Fernandez Brusco A, Torrecillas Camacho C, Villabona Lopez E, Rodriguez Belloso M, Villares Urgell L, Serrano Barrena P, Morán Bueno N
Integrating Extended Reality Into Primary Care Chronic Pain Programs via the REDOCVR Intervention: Real-World Implementation Feasibility and Usability Study
JMIR XR Spatial Comput 2025;2:e82858
URL: https://xr.jmir.org/2025/1/e82858
doi:10.2196/82858

Association Between Sensation Seeking and Fear Response: Interventional Study of Personality and Behavior Using a Virtual Reality Heights Simulation

Rebecca C Daugherty[1], BSc; Christina S Jacksack[1,2], BSc; Becca J Daye[1], BSc; Brandon G Oberlin[1,2], PhD

[1]Department of Psychology, School of Science, Indiana University Indianapolis, 420 University Blvd, Indianapolis, IN, United States

[2]Department of Psychiatry, Indiana University School of Medicine, 355 W 16th Street, Indianapolis, IN, United States

Corresponding Author:
Brandon G Oberlin, PhD
Department of Psychology, School of Science, Indiana University Indianapolis, 420 University Blvd, Indianapolis, IN, United States

Abstract

Background: Immersive virtual reality (VR) technology presents digital simulations that create the sense of an actual experience. VR simulations are persuasive enough to elicit physiological reactions that mirror real-world responses. Prior research suggests that fear responses and sensation seeking are inversely correlated, but that work largely relies on self-reported outcomes and hypothetical scenarios.

Objective: To more closely model real-world phenomena, we tested for inverse associations using an experiential height exposure simulation and a behavioral task for sensation seeking. We tested these associations comprehensively by using multiple methods for convergent evidence.

Methods: A total of 57 healthy undergraduates participated in an interventional study that included an anxiety-inducing VR simulation, behavioral tasks, and personality inventories. The VR paradigm (Richie's Plank) prompted users to walk across and step off a plank at the top of a skyscraper. This simulation of extreme height exposure and falling was intended to evoke fear. Physiological recordings and self-reported state anxiety were collected prior to and during the experience. Behavioral sensation seeking was quantified using an olfactory choice task offering a "boring" or "exciting" (risky) option varying in intensity and pleasantness. Evoked fear was calculated as the difference between the calm (pre-plank) and provoked fear state (standing on the plank), with correlations performed between evoked fear and personality and behavioral measures. The false discovery rate was set to $q<.05$, with analyses conducted in SPSS.

Results: The VR experience evoked self-reported fear ($P<.001$) and physiological arousal ($P<.006$ for heart rate and $P<.017$ for respiration). Acrophobia correlated with self-reported fear in men and women ($P<.01$). Behavioral sensation seeking negatively correlated with both self-reported fear ($P=.02$) and increased heart rate in men ($P=.02$). Behavioral and self-reported sensation seeking were uncorrelated ($P=.89$). Self-reported fear was uncorrelated with physiological fear responses ($P>.46$).

Conclusions: VR simulations can produce lifelike responses to scenarios that are impractical to test in reality. Our demonstration of an experiential manipulation negatively correlating with sensation seeking behavior in men increases confidence in other findings from studies using more traditional methods. Our use of VR along with objective measures, for example, behavioral tasks, and subjective measures, for example, self-report, confirm the effectiveness of these tools to investigate behavioral health topics. Our findings further suggest that sex is an important intervening factor for fear and sensation seeking and that additional study is warranted. This study highlights the potential of VR to expand convergent validity more broadly with other traits and paradigms. Finally, VR technology permits presenting highly abstract or improbable scenarios, thus expanding the range of topics for behavioral investigations. Given the ever-wider adoption of immersive therapeutics in the clinic, VR research will continue to facilitate the study of biobehavioral outcomes and interactions with personality factors and advance actionable knowledge for clinical applications.

(JMIR XR Spatial Comput 2025;2:e73785) doi:10.2196/73785

KEYWORDS

behavioral psychology; immersion; immersive virtual reality; novelty seeking; real-life responses; phobia; risk-taking; simulation

Introduction

Immersive virtual reality (VR) produces the functional and behavioral equivalence of actually being in a place, known as "presence" [1-5]. Immersion arises from high-fidelity stereoscopic visual and 3D audio stimuli that respond to user inputs; together, these create an "inclusive, extensive, surrounding, and vivid illusion of reality" [6]. Germane to clinical research, VR can deliver experiences that feel authentic but would otherwise be too dangerous, costly, or impractical to administer in a laboratory setting [1]. Until recently, research on evoked fear responses generally relied on affective pictures, video, and threat scenarios [7-9]. Extending prior research demonstrating that increasing immersive properties increased evoked responses [10,11], VR technology offers the promise of delivering experiences that are interactive, vivid, powerful, customizable, and standardized for studying clinically relevant outcomes.

Fear is an arousal response to threat stimuli, whereas anxiety occurs while approaching or anticipating a threat [12]; however, the distinction between fear and anxiety becomes blurred as the threat gets closer or more certain [13]. Unreasonable fear and anxiety elicited by particular stimuli can manifest as specific phobias, which occur in ~10% of the population and are highly comorbid, especially with anxiety, mood, and personality disorders [14]. The fear of heights is the second most common among these specific fears, with a prevalence of 4.5% [14], and is twice as common in women than in men [14,15]. The physiological fear response is mediated by the Fight/Flight/Freezing System and autonomic arousal, which can be readily quantified using altered heart rate [12,16]. Longitudinal evidence suggests that evoked physiological responses are more reliable than baseline physiological measures [17]. Further, physiological responses appeared to be more sensitive than self-reported valence to simulated threats when those threats were more lifelike and created by computer-generated graphics [11]. Immersive VR compared to 2D versions of provocative scenes evoked stronger physiological and emotional responses, coinciding with a stronger sense of presence [10,18]. Prior work indicates that virtual height simulations effectively elicit fear-related arousal [5,19]. We reasoned that a realistic experiential fear provocation would be highly appropriate for investigating associations with other objectively measured behavioral outcomes.

Sensation seeking, defined as "the seeking of varied, novel, complex, and intense sensations and experiences, and the willingness to take physical, social, legal, and financial risks for the sake of such experience" [20], is a putative stress buffer [21]. It is negatively associated with anxiety [19,22,23], risk perception [22,24], and fear responses [25]; however, the associations differ by sex (particularly on the Experience Seeking dimension) [26]. Unsurprisingly, fear-inducing activities (eg, hang-gliding, BASE jumping) attract high sensation seekers [27,28]. Moreover, many "extreme sports" (eg, bungee jumping, skydiving, ultralight piloting) evoke the fear of falling, a powerful and common "natural fear" in humans [29].

Sensation seeking is traditionally measured with self-report inventories [30,31]. To quantify sensation seeking and overcome some of self-report's intrinsic limitations, such as social desirability, self-awareness, cultural limitations, and criterion contamination [32-35], we created a behavioral task that models the key aspects of this trait [36]. The Aroma Choice Task (ACT) quantifies sensation seeking behavior as a binary choice between olfactory stimuli varying in intensity, novelty, and riskiness using actual sensory experiences presented in real time. Sensation-seeking behavior in this task correlates with self-reported sensation seeking [36], reward-related brain activation [37], and alcohol-induced shifts in reward preference [38]. Given the body of work suggesting that studies relying on low-realism tasks, hypothetical choices, and self-report questionnaires limit neurobehavioral [39] and behavioral genetic [40] conclusions, our objective sensation seeking task quantifies the behavioral trait as described by Zuckerman [30].

While virtual height simulations consistently elicit physiological reactions [5,19,41,42], only a few studies have examined relationships between behavior and sensation seeking; however, in VR [19,43], no studies, to the best of our knowledge, have tested for associations of physiological fear responses and behavioral sensation seeking. We expect more ecologically valid manipulations (ie, greater resemblance to real-world experience) of fear and anxiety, such as a realistic VR simulation of extreme heights, paired with objective behavioral measures of sensation seeking to reveal more generalizable findings for clinical research. Here, we elicit fear responses with an immersive interactive height exposure simulation [44] and test for associations with sensation seeking behavior. We quantify fear responses as subjective (evoked state anxiety) and objective (evoked physiological arousal). We hypothesize that (1) the virtual height simulation will increase evoked state anxiety and physiological arousal, (2) evoked state anxiety and physiological fear responses will be positively associated, (3) height anxiety will correlate with evoked state anxiety, (4) behavioral sensation seeking and self-reported evoked state anxiety will be negatively associated, and (5) behavioral sensation seeking and physiological fear response will be negatively associated. We also explore potential correlations between behavioral and self-reported sensation seeking.

Methods

Ethical Considerations

The Indiana University Institutional Review Board approved all recruiting and study procedures (protocol number 17398). Students provided informed consent before study participation. Privacy and confidentiality were maintained for all participants, and compensation was awarded in the form of class credit.

Participants

A total of 57 healthy undergraduate students were recruited from an urban midwestern university via online listings through the university's Sona Systems human subjects pool; course credit was provided for participation. Students provided informed consent before study participation. Exclusions included poor sense of smell, extreme sensitivity to odors or volatile chemicals, chronic or current asthma, pregnancy or nursing, or

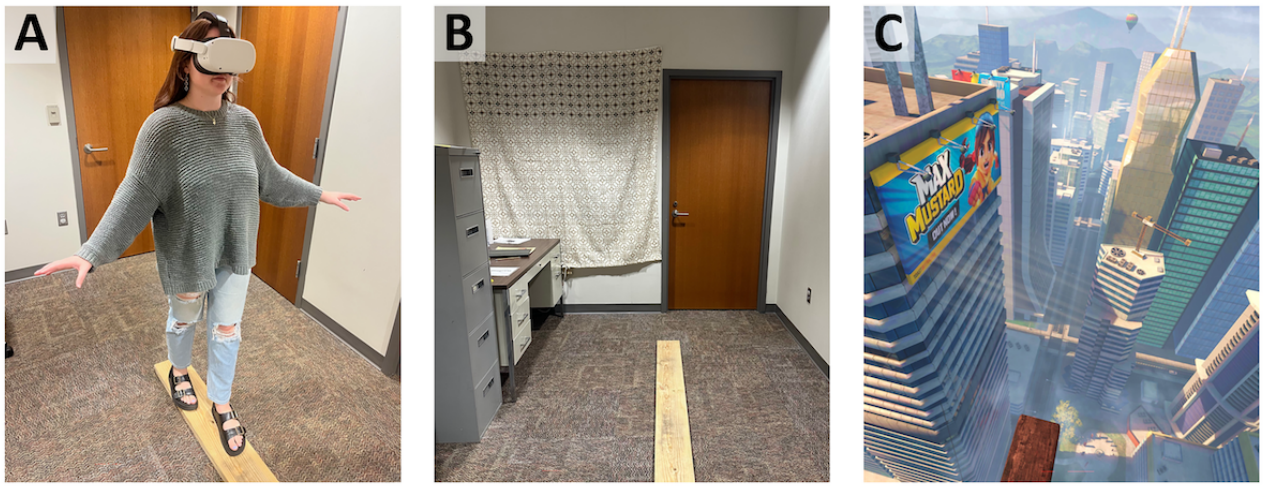
the use of a nasally administered medication (excepting steroids). Physiological data collection devices were unavailable in the beginning of the study, so those data were not collected from 21 participants. Technical challenges additionally limited data collection for heart rate and respiration (n=1), heart rate only (n=2), and respiration only (n=1). At least 1 participant did not feel well enough to complete the study and provided only demographic and personality data.

Procedure

Prior to the VR experience, participants provided demographic information and self-reports of anxiety, height-specific anxiety (major component of acrophobia), and impulsive sensation seeking. All self-report inventories were administered via Qualtrics. Participants also completed the behavioral sensation-seeking task. Participants were then guided through an immersive fear-inducing VR simulation of height exposure, with physiological recording initiated before the VR experience and continuing until the VR task was completed. Self-reported state anxiety was measured again near the end of the fear induction while in VR. Figure 1 illustrates the experimental procedures in a timeline format.

Figure 1. Experimental procedures. Undergraduate student participants were tested in the same manner in a pre-post intervention design. Intake included screening and informed consent, followed by a demographics questionnaire (eg, age, sex, childhood income, race). All procedures were completed within 60 minutes and were conducted in the same experiment room (pictured in Figure 2A and B). Behavioral sensation seeking was assessed with the Aroma Choice Task (ACT). Personality questionnaires included the State-Trait Anxiety Inventory for States (STAI-S), Acrophobia Questionnaire-Anxiety, and Zuckerman Kuhlman Personality Questionnaire (ZKPQ). The Richie's Plank Experience is described in detail in the text. Physiological data (heart rate and respiration) were recorded throughout virtual reality (VR).

Figure 2. Experimental environment and display. (**A**) All experimental procedures were conducted in the private quiet room shown, with participants' initial steps corresponding to the tactile experience of walking on a plank. (**B**) The researchers' view of the room and plank, and (**C**) the visual effects as displayed to participants. Wind noise corresponding to high altitudes was played through the headset speakers while walking on the plank. The tactile, visual, and auditory elements integrated for an immersive visceral simulation of walking on a plank from a city building at extreme height. Participants were oriented facing the plank at the onset of the paradigm.



Self-Report Inventories

Anxiety

Changes in self-reported anxiety before and during the experience indexed evoked fear. The State-Trait Anxiety Inventory for states (STAI-S) is a 20-item questionnaire rated on a Likert-type scale anchored by "not at all" (1) to "very much so" (4). Participants report how they currently feel (eg, *"I feel frightened"*). Total scores range from 20 to 80, with 80 indicating the highest anxiety levels [45]. This questionnaire assessed baseline and evoked anxiety, with the subtracted (Pre-Plank vs Plank) difference quantifying evoked anxiety. We changed the administration of the STAI-S during data collection to better capture the emotional state while still in VR. The first 21 participants completed the inventory on a laptop following the experience, but the last 36 participants were queried verbally while still in the headset (standing on the end of the virtual plank), and responses were recorded by the experimenter. "Plank," the post condition, refers to both ways of measurement. This change was provoked by the concern based on behavioral observation that STAI-S responses following the experience might be substantially influenced by relief, that is, anxiety alleviation, upon the termination of the fear-inducing experience. To ensure consistency, both the Pre-Plank and Plank STAI-S inventories were conducted verbally following this change.

Height-Specific Anxiety

The Acrophobia Questionnaire-Anxiety is a 20-item questionnaire that poses hypothetical height-related fear scenarios (eg, "*looking down a stairway from several flights up*") and collects responses on a 7-point Likert-type scale anchored by "not at all anxious; calm and relaxed" (0) to "extremely anxious" (6). Total scores ranged from 0 to 120 with higher scores indicating greater levels of anxiety specifically related to heights, or acrophobia [46].

Impulsivity and Sensation Seeking

The Zuckerman Kuhlman Personality Questionnaire is a 50-item forced-choice inventory posing self-descriptive statements (eg, "*I often do things on impulse*"). The 5 subscales are impulsive sensation seeking, neuroticism-anxiety, aggression-hostility, activity, and sociability. Possible scores on each ranged from 1 to 10, with 10 indicating a high presence of the trait [47].

Behavioral Sensation Seeking

The Aroma Choice Task (ACT) is a validated behavioral test of sensation seeking that measures the relative preference for an intense, novel, varied, risky option versus a mild, safe, "boring" option, with odorants delivered in real time. Participants are instructed:

> *For the next 12 minutes, you will make choices about some smells. The choice labeled 'Standard' will likely be mild and pleasant. The choice labeled 'Varied' will likely be stronger and pleasant, but there is a chance that it will be unpleasant. Upon making a choice, please inhale deeply through your nose to receive the aroma.*

Choice ratio, the percentage of "Varied" choices out of a total of 20 binary choice trials (range: 0% - 100%), yields a single behavioral index reflecting behavioral sensation seeking (designed after self-reported sensation-seeking trait descriptions) [30,31]. The original ACT was developed with an air dilution olfactometer [36], but a simpler, manual version yielded analogous results [38]. We further modified the task to deliver 20 trials instead of the original 40, as our prior work indicated that the first 20 trials accurately capture the trait with lower participant burden [36].

Physiological Recording

Heart rate, respiration, and skin conductance were collected for 36 participants using the BioRadio (Great Lakes Neurotechnologies; Cleveland, OH, USA) and skin electrodes plus a respiratory inductance plethysmography belt, with data logged on a laptop computer. At least 1 researcher closely assisted the participant during VR to prevent collisions and stumbles. A second researcher entered live event markers (single keystrokes) that were logged in the BioRadio data stream as the participant progressed through the phases of the VR experience. Standardized breathing exercises (eg, Balban et al [48]) prior to VR were intended to mitigate physiological variability between participants. The "Pre-Plank" measurements included the anticipatory elevator ride up to altitude and the 20 seconds before stepping onto the plank (mean 0.74 [0.24] minutes; max 1.70). The "Plank" measurements comprised stepping onto the plank after the 20 seconds was complete and then walking the length of the plank and off into freefall (mean 1.88 [0.60] minutes; max 2.85). Skin conductance data were not analyzed.

Virtual Reality

The fear of falling from heights is an "innate" fear, that is, nonreliant on associative conditioning [29] or locomotion experience [49], and nearly universal [50], meaning it is highly generalizable and reliable for eliciting potent fear responses. An immersive height exposure simulation was delivered on the Meta Quest 2 head-mounted VR display. The Meta Quest 2 features 6 degrees of freedom, high refresh rate (up to 120 Hz), and adjustment for interpupillary distance [51,52]. These features, combined with limited movement and short VR duration (mean 2.90 [0.58] minutes, max 3.90), minimized the risk of VR sickness [53]. Nonetheless, participants were informed about the risk of VR sickness and were encouraged to notify the researcher at any time to end the simulation. No participants complained of VR sickness or feeling unwell due to the VR experience.

The immersive height exposure simulation was the "Richie's Plank Experience" [44] delivered with a Meta Quest 2 head-mounted VR display. This simulation is widely used in VR research on fear responses and behavior [43,54,55]. It has also been used to study risk-taking [56] and suicide willingness [57] and is even proposed as a psychological preparation for psychedelic experience [58]. The paradigm simulates an elevator ride to the top of a tall building, and upon the door opening, presents a cityscape view from heights (~80 stories). Participants walk onto, down the length of, and off the end of a wooden plank protruding from the skyscraper. The height illusion is conveyed by audio and visual simulation of extreme exposure to open space and reified with wind noise and birds flying below the participant. Immersion was maximized with haptic feedback from a real wooden plank (6' long "2×8" [2 m×4 cm×19 cm]) spatially registered to the virtual plank. The wooden plank was slightly warped and thus creaked and shifted with human weight. Although verbal instructions were provided as needed, researchers refrained from unnecessary verbal or physical interaction during the experience to preserve presence. A simulated participant (the first author) walks on the plank (Figure 2A) in the center of partitioned office space (Figure 2B), with the participants' view at heights illustrated in Figure 2C.

Analytical Strategy

Physiological data collected by the BioCapture (Great Lakes Neurotechnologies; Cleveland, OH, USA) software were modeled in VivoSense (version 3.4). VivoSense calculated the means of heart rate and respiration rates in 10-second bins. Changes in physiological measures were calculated as the average rate during anticipation of stepping on the plank (Plank) minus the baseline average (Pre-Plank). All statistical tests were performed in SPSS v29.0.2.0 (IBM). Analyses were stratified by sex as prior work demonstrated important interactions by sex between anxiety and sensation seeking [26] and by sex and elements of sensation seeking [30,59]. VR-induced increases in anxiety and physiological arousal were tested with paired *t* tests (2-tailed) between baseline "Pre-Plank" and the moment

of stepping off the plank "Plank." Change scores were calculated for self-reported state anxiety and physiological measures (Plank minus Pre-Plank scores). These were tested for correlations with behavioral sensation seeking, each other, and height-specific anxiety. The potential effect of changing the STAI-S assessment method (computer vs verbal inside VR) was tested with independent-samples *t* tests (2-tailed). We tested for differences in participant characteristics of those with physiological recording versus without to assess potential effects on interpretation. *t* tests (2-tailed) of inhomogeneous variance (Levene test) were performed using the Welch *t* test (2-tailed). The false discovery rate was controlled with the Benjamini-Hochberg procedure [60], and limited to 5% (q<.05), by setting α=.0278 for individual tests in a priori hypotheses.

Results

Demographics and Personality

Tests evaluating baseline sex differences (χ^2 for nominal, *t* test [2-tailed] for continuous) reported were uncorrected and descriptive in nature. All in-text data are reported as mean (SD). Women reported higher baseline levels of anxiety than men did (neuroticism-anxiety, state anxiety, and height-related anxiety, P<.03; Table 1). No other sex differences were detected.

Table . Participant characteristics: demographics and personality (n=57).

Characteristic	Women (n=40)	Men (n=17)
Age (y), mean (SD)	20.23 (2.94)	20.06 (2.19)
Childhood income[a] (US $), median (IQR)	86,000 (35,000-127,000)	86k (35,000-141,000)
Race, n (%)		
American Indian	1 (3)	0 (0)
Asian	6 (15)	1 (6)
Black	6 (15)	2 (12)
Other/Unknown[b]	6 (15)	2 (12)
White	21 (53)	12 (71)
ZKPQ[c], mean (SD)		
Impulsive sensation seeking	3.82 (2.41)	4.35 (2.57)
Neuroticism-anxiety	5.50 (3.11)[d]	2.06 (2.08)
Aggression-hostility	4.24 (1.84)	3.88 (2.20)
Activity	5.05 (2.68)	4.41 (3.24)
Sociability	4.05 (2.52)	3.65 (3.18)
STAI-S[e], mean (SD)	34.26 (9.31)[f]	28.59 (7.87)
Acrophobia questionnaire-anxiety, mean (SD)	55.60 (18.78)[g]	43.12 (19.69)

[a]Reported as median and interquartile ranges representing the geometric means of the inventory ranges (<US $10,000, $10,000-$24,000, $25,000-$49,000, $50,000-$74,000, $75,000-$99,000, $100,000-$199,000, $200,000-$500,000, >$500,000).

[b]Hispanic ethnicity: n=5 identified as Other/Unknown, n=3 as White, n=1 as American Indian. Hispanic ethnicity did not differ by sex or race (χ^2, P>.71)

[c]ZKPQ: Zuckerman Kuhlman Personality Questionnaire subscales.

[d]$t_{44.7}$=4.83, P<.001.

[e]STAI-S: State Trait Anxiety, Inventory for states.

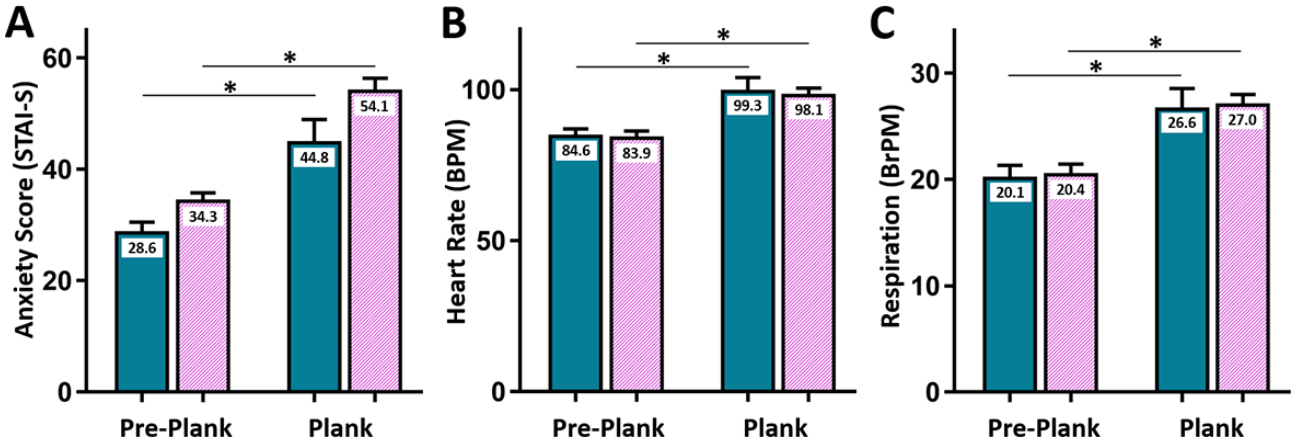
[f]t_{54}=2.19, P=.03.

[g]t_{55}=2.26, P=.03.

Evoked Fear: Increased Anxiety

The VR experience increased anxiety in both men (t_{16}=5.29, P<.001, 28.59 [7.87], and 44.76 [17.27], q<.05, Pre-Plank and Plank, respectively) and women (t_{38}=9.50, P<.001, q<.05, 34.26 [9.31] and 54.05 [14.51]; Figure 3A). The increase in anxiety was not significantly different before versus after the data collection method change (for more details, see the Anxiety section; P=.75).

Figure 3. Effects of VR. Participants (solid cyan bars show the values for men, white/violet diagonal hash bars show the values for women) showed increased (**A**) anxiety states, (**B**) heart rate, and (**C**) respiration when on the plank relative to the Pre-Plank baseline. STAI-S=State trait anxiety inventory, State; BPM=beats per minute; BrPM=breaths per minute; *q<.05.



Physiological Arousal

In the 36 participants from whom physiological data were collected, the VR experience increased heart rate in both sexes (t_8=3.84, P=.005 for men t_{23}=6.99, P<.001 for women; q<.05; Figure 3B). Similarly, the VR experience increased respiration in both sexes (t_8=3.06, P=.02 for men and t_{24}=5.35, P<.001 for women; q<.05; Figure 3C). Of note, characteristics (Table 1) did not differ between participants with physiological data collected versus those without, q>.05 (11 tests); P=.53, P=.046, P=.24, P=.51, P=.84, P=.88, P=.62, P=.60, P=.25, P=.50, and P=.61, corresponding to childhood income, age, impulsive sensation seeking, aggression-hostility, activity, sociability, neuroticism-anxiety, AQ-anxiety, STAI, sex, and race, respectively.

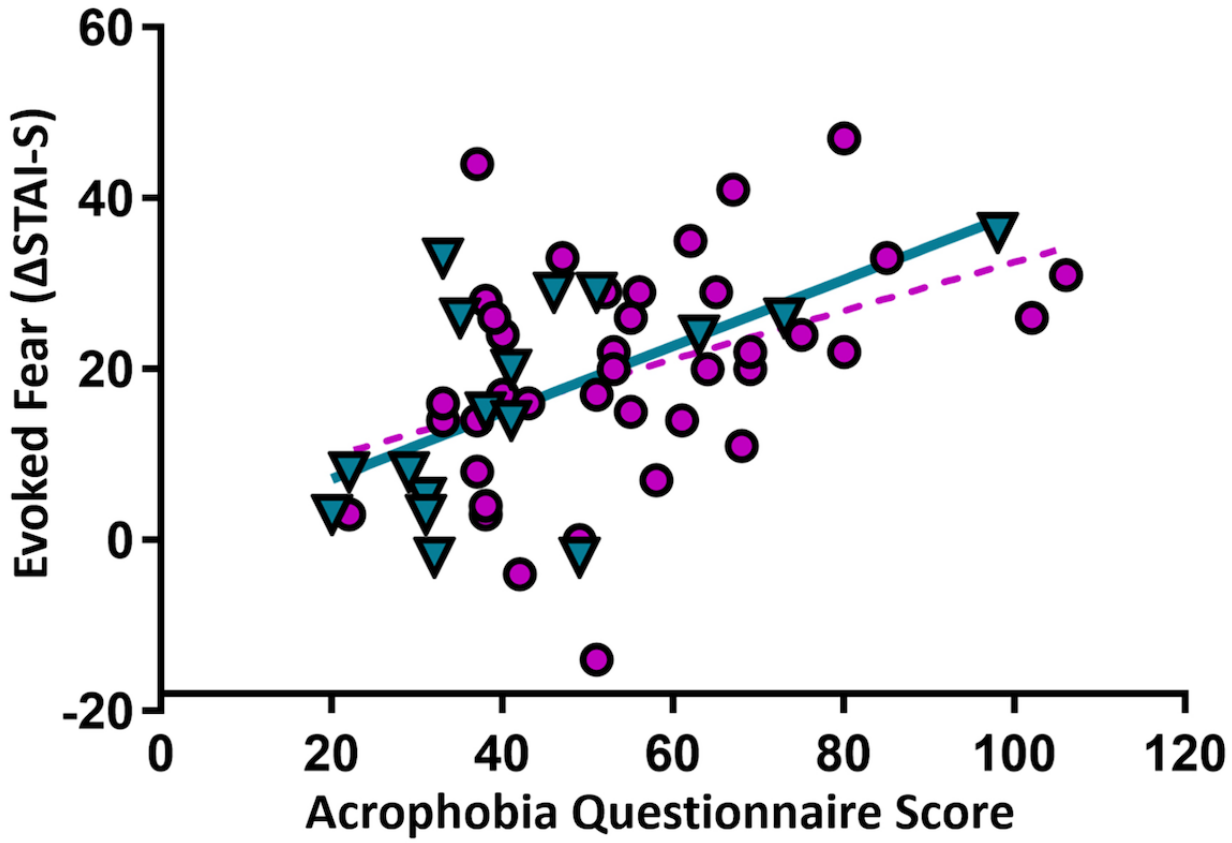
Evoked Fear and Physiological Response

Fear (evoked state anxiety) did not correlate with changes in heart rate (P=.60 and .79 for men and women, respectively) nor respiration (P=.47 and .48) in the participants from whom physiological data were collected. Baseline anxiety was also uncorrelated with changes in heart rate (P=.70 and .29) and changes in respiration (P=.71 and .17).

Acrophobia and Evoked Fear

Fear of heights (acrophobia) was correlated with evoked state anxiety in men (r(15)=.606, P=.01) and women (r(38)=.410, P=.009, q<.05; Figure 4).

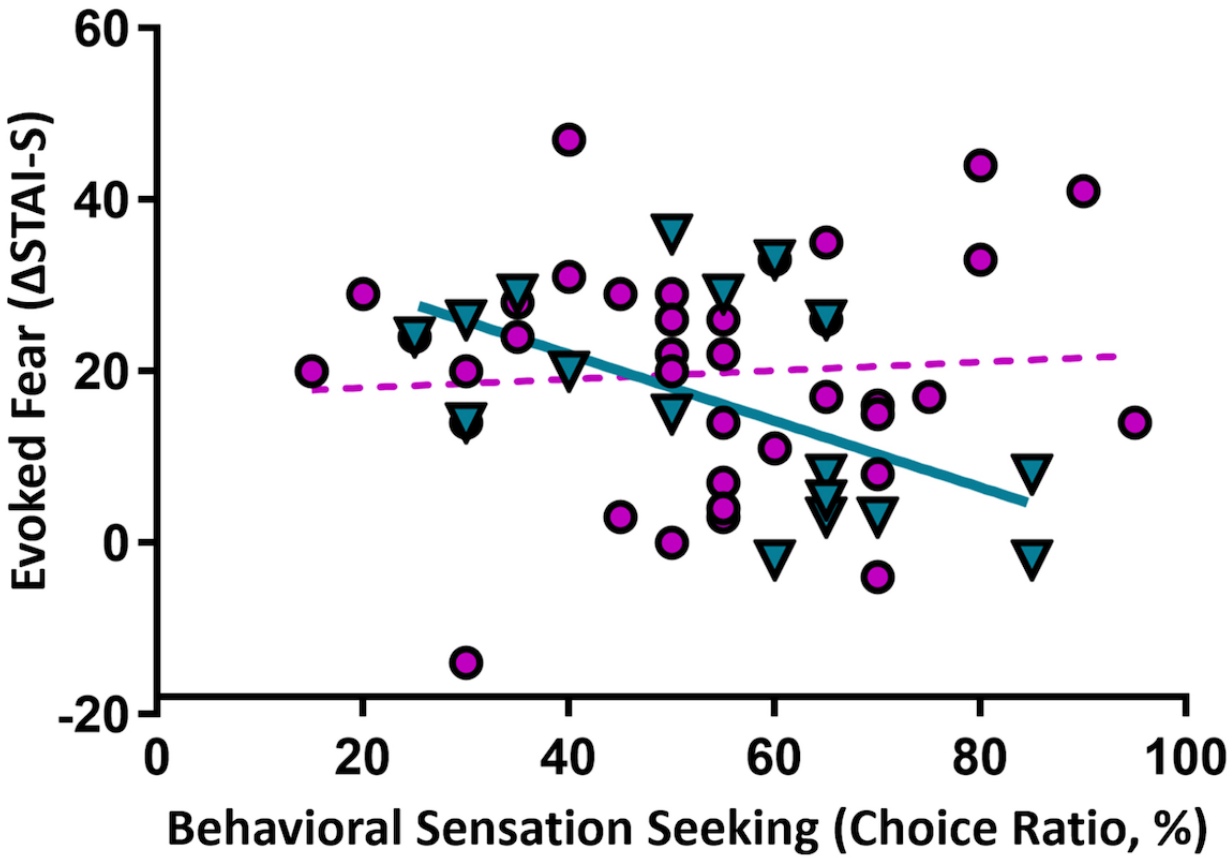
Figure 4. Height-related fear and fear response. Higher scores on the Acrophobia Questionnaire in both men (cyan triangles, solid line) and women (violet circles, dotted line) predicted greater anxiety reactivity to the virtual reality (VR) experience (r>.40, P<.011, q<.05). Regression lines illustrate strength and direction of associations. ΔSTAI-S: Plank minus Pre-Plank STAI-S scores.



Behavioral Sensation Seeking and Evoked Fear

High intensity preference (ACT scores) negatively correlated with evoked state anxiety (Δ STAI-S) in men (r(15)=−.559, P=.02, q<.05), but not in women (P=.67; Figure 5). Preference for high intensity was not correlated with baseline STAI-S scores in women (P=.61), although there was a trend in men (r(15)=−.465, P=.06).

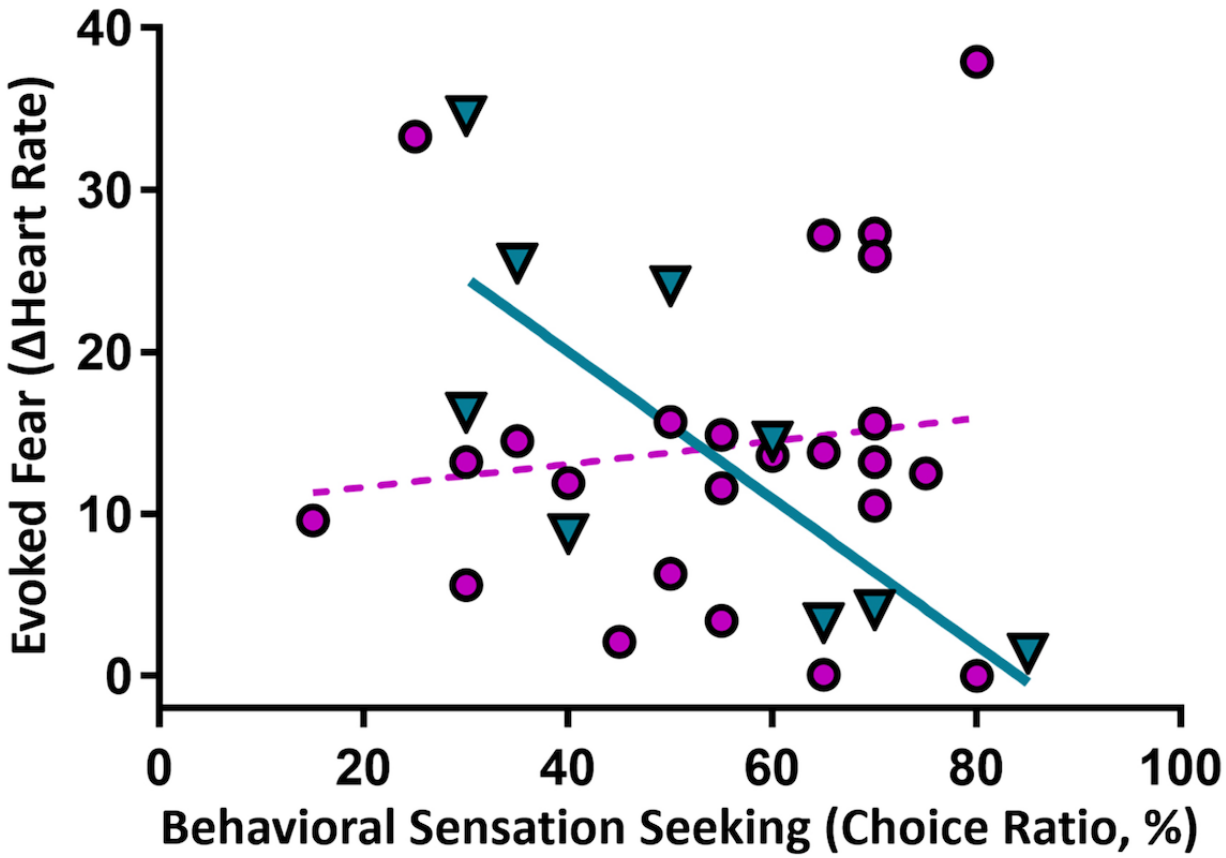
Figure 5. Sensation seeking and fear induction. Higher sensation seeking in men (cyan triangles, solid line) predicted lower anxiety response in virtual reality (VR*; r*=−.559, *P*=.02, *q*<.05), but not in women (violet circles, dotted line), *P*=.67. Regression lines illustrate strength and direction of associations. Choice ratio is the percentage of "varied" choices selected during the aroma choice task. ΔSTAI-S: Plank minus Pre-Plank STAI-S scores.



Behavioral Sensation Seeking and Physiological Fear Response

High intensity preference (ACT scores) negatively correlated with increased heart rate in men (*r*(7)=−.771, *P*=.02, *q*<.05), but not women (*P*=.54; Figure 6). Increased respiration was uncorrelated in both sexes (*P*>.28).

Figure 6. Sensation seeking and physiological fear response. Higher sensation seeking in men (cyan triangles, solid line) predicted lower heart rate reactivity to the virtual reality (VR) experience (r=−.771, P=.02, q<.05), but not in women (violet circles, dotted line), P=.54. Regression lines illustrate strength and direction of associations. Choice ratio is the percentage of "Varied" choices selected during the aroma choice task. ΔHeart Rate: Plank minus Pre-Plank heart rate measurements.



Behavioral and Self-Reported Sensation Seeking

High intensity preference did not correlate with self-reported sensation seeking in men (P=.47) or women (P=.72); collapsing across sex did not permit detection of an association (P=.89). See Table 2 for correlations of personality and behavioral metrics (uncorrected).

Table . Systematic matrix of correlations.

	1	2	3	4	5	6	7	8	9	10
ZKPQ[a]										
1. ImpSS	—	0.082	0.286[b]	0.324[b]	−0.036	0.019	−0.173	−0.224	−0.027	0.212
2. Agg-Host		—	0.104	0.163	0.197	0.106	0.087	−0.134	0.238	0.349[b]
3. Act			—	0.202	0.210	−0.130	0.187	0.159	0.354[b]	0.127
4. Sy				—	−0.001	0.073	−0.097	0.059	0.229	0.261
5. N-Anx					—	−0.140	−0.332[b]	0.309[b]	−0.073	−0.142
Beh. SS (ACT[c])										
6. Choice ratio						—	−0.111	−0.236	−0.012	−0.029
Reported anxiety										
7. ΔS-TAI-S							—	0.483[d]	−0.039	−0.034
8. AQ-Anx								—	−0.318	−0.128
VR physiology										
9. HR Δ									—	0.137
10. Resp Δ										—

[a]ZKPQ: Zuckerman Kuhlman Personality Questionnaire subscales.

[b]Correlation is significant at $P<.05$.

[c]ACT: Aroma Choice Task.

[d]Correlation is significant at $P<.01$.

Discussion

Primary Findings

We found support for our hypotheses concerning VR's capacity to evoke fear (self-reported changes in state anxiety and physiological arousal) and the association between evoked fear and fear of heights. We found support for the negative association between evoked fear and behavioral sensation seeking in men, but not women. This unexpected finding is potentially due to higher trait anxiety scores in women, that is, imposing a ceiling effect and making associations between elicited anxiety and other traits difficult to detect. Interestingly, evoked fear measures were uncorrelated. Behavioral and self-reported sensation seeking were uncorrelated. These findings suggest that VR experiences possess sufficient ecological validity to elicit subjective and objective fear responses mirroring responses to real-world scenarios. Moreover, responses to the digital experience reflect associations with other traits (acrophobia and sensation seeking).

This study demonstrates the potential of VR for neuroscience and clinical research. VR continues to offer considerable promise for clinical and research applications alike. The expanding reach of VR is facilitated by better immersion technology, decreasing cost, creative applications, and wider adoption. While VR has long been used to administer exposure therapy [61-63] and treat pain [64-66], emergent applications target increasingly abstract constructs [67,68]. Germane to this study, VR applications effectively treat various anxiety disorders (eg, specific phobias, social anxiety disorder, panic disorder) in randomized controlled trials, producing effects comparable to conventional treatment and significantly better than passive controls [69]. In addition to promising efficacy data, providers and patients appear eager to adopt VR methods, as suggested by 2 recent reports on integrating VR in clinical practice [70,71]. The immersive nature of this nascent technology permits new avenues of investigation and permits research on humans that would be dangerous or impractical to study with real stimuli. We believe that the potential of VR to unite human and animal paradigms heralds a new translational era wherein strictly controlled animal neuroscience experiments can be accurately replicated in humans. For example, the elevated plus-maze—the gold standard in rodent anxiety research—can be instantiated for human participants [19] to connect basic neuroscience in rodents with human behavioral data. Other behavioral paradigms widely used in rodents, such as conditioned place preference, can now be accurately reproduced in humans using VR [72], producing convergent findings in both research contexts. Thus,

the role and relevance of VR for both laboratory research and clinical practice are expected to grow substantially in the near future. The use of VR permits testing extant theoretical knowledge in more lifelike settings and experiences, and when paired with behavioral tests, yields increasingly objective outcomes.

VR can add knowledge to mature bodies of research, such as approach-avoidance, by presenting realistic simulations in humans. Approach-avoidance describes behavior that orients organisms toward positive and away from negative stimuli, respectively [73]. Approach-avoidance tendencies are likely rooted in evolutionary factors, primarily through sexually divergent selection pressure [74]; that is, reproductive fitness optimized by exploratory behaviors in men [75,76] and harm avoidance in women [77,78]. The extremes of the approach-avoidance continuum are marked by exaggerated attention on reward or threat cues [79]. These tendencies emerge from overactive brain reward or motivational systems [80] and underregulated brain threat systems [81] for approach, and conversely, overactive brain threat systems [82] and underactive brain reward systems [83] for avoidance. Sensation seeking and fear represent aspects of approach and avoidance, that is, opposing processes that modulate threat responses, such that high sensation seekers are less physiologically responsive to threat stimuli than low sensation seekers. One study testing fear responses found that high sensation seekers showed no response to threatening stimuli (versus control stimuli), whereas low sensation seekers produced an 8-fold increase in electromyographic response to threats [25]. No group difference in self-reported emotional reactivity was detected, further supporting the value of objective measurements.

Men are higher sensation seekers than women (particularly thrill or adventure seeking and disinhibition) [84,85], but the relationship between sensation seeking and fear appears to differ by sex. Investigating this relationship as an interaction with sex, Blankstein [26] found that the Sensation Seeking Scale (SSS) [23] total score negatively correlated with anxiety reactivity (Activity Preference Questionnaire) total and subscales (Social and Physical) at $r>.43$, $P<.01$ in men, but not in women ($r<.07$). However, a similar study found a number of negative correlations between the Sensation Seeking Scale and anxiety-related items (S-R Inventory) in both sexes [86], indicating mixed results in detecting sex interactions with approach-avoidance correlations. The lack of consilience in prior work might be explained by either (1) dependence on self-report inventories, with self-reported fear and sensation seeking often incongruent with objective measures [38,87], or (2) high anxiety and low sensation seeking in women producing restricted ranges (ceiling and floor effects), making correlations difficult to detect. Both factors potentially contribute together to the divergence. Future well-powered studies, ideally using precise behavioral tasks, should clarify these possible associations.

We did not detect correlations between self-reported fear and physiological responses to height exposure. While this is perhaps a surprising result, prior work suggests that self-reported fear does not necessarily reflect biological responses. In a real-world test, participants' self-reported fear of crime did not differ between walking down a dimly-lit path (vs well-lit control), but the dimly-lit path participants' heart rate increased by 17% ($P=.002$), with that in the controls remaining unchanged [87]. Even patients with anxiety disorders do not accurately report the degree of physiological responses to stress in laboratory tests [88,89]. This lack of concordance may be explained by individual differences in interoceptive ability [90]. The disconnection between self-reported traits and objective measures is also found in behavioral assessments of impulsivity [91], empathy [92], and risk preference [93], suggesting that the incongruence extends well beyond fear and anxiety. A recent report on associations between interoceptive ability and autobiographical memory [94] indicates that interoceptive perception (physical self-awareness) relates to episodic recall (cognitive self-awareness) and suggests the intriguing possibility that individual differences in these domains may be governed by some larger self-awareness meta factor.

Behavioral seeking and self-reported sensation seeking were uncorrelated in this sample. Existing findings offer scant data on this association due to the absence of behavioral sensation seeking tasks. Related traits such as impulsivity and risk-taking reveal low agreement between behavioral and self-reported assessment; for example, various measures of impulsivity are correlated at $r=\sim0.1$ in meta-analysis [91], and practically no relationship is observed between risk-taking measures [95]. Low reliability is observed in behavioral task data across domains [96]; in parallel, self-reported data suffer from various serious forms of bias [97]. The "jingle fallacy" (conflating interpretation of 2 measures because they have the same name) [98] exacerbates this problem. While the lack of convergence between behavioral and self-reported findings is perhaps unintuitive, this divergence between self-reported and behavioral data represents opportunities to discover additional features of personality traits. That is, important features of a given trait may not be fully captured by any single task or inventory.

Limitations

Some limitations should be acknowledged. First, the sample would benefit from more power. The homogeneity of the sample—reflecting typical undergraduates—is predominately female and White, precluding well-powered direct comparisons by sex and potentially limiting generalizability. Generalizability would be enhanced by a community sample with a more even sex distribution and a larger range of age and socioeconomic status. Finally, the truncated sample of subjects providing physiological data was suboptimal, as physiological data were only collected from 36 of the 57 participants, although the absence of significant differences between the subsamples somewhat mitigates this concern.

Conclusions

The current report demonstrates the potential use of VR for neuroscience and clinical research. Beyond research and education, VR is now well established as a clinically valuable tool in health care. The combination of using VR, objective measures (physiological recordings and behavioral tasks), and subjective measures (self-report) to investigate a behavioral health topic allows more rigorous investigation than any one of these approaches alone. Through these measures, we confirmed

associations between self-reported experience and physiological fear in response to heights, in addition to behavioral patterns and personality related to sensation seeking. Further evidence of the disconnection between objective and self-reported methods was found, although this was perhaps unsurprising. We expect ever-wider adoption of VR applications and objective measures in the clinic and continued expansion in the laboratory research domain.

Acknowledgments

We wish to thank Dr. Melissa Cyders for generous use of her laboratory space, and additionally, we wish to thank Sampada Bhatnagar and Sarah Turo for their help with training, preparing the experiment, and technical help with software. No artificial intelligence tools (LLMs, eg, ChatGPT) were involved with any stage of ideation or manuscript preparation. No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BGO, BJD, CSJ, RCD
Formal analysis: BGO, CSJ, RCD
Investigation: BGO, CSJ, RCD
Methodology: BGO, BJD
Project administration: BGO, CSJ, RCD
Validation: BGO, CSJ, RCD
Visualization: BGO, CSJ, RCD
Writing – original draft: BGO, BJD, CSJ, RCD
Writing – review & editing: BGO, BJD, CSJ, RCD

Conflicts of Interest

None declared.

References

1. Bailenson JN. Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do, 1st edition: W.W. Norton & Company; 2018. URL: https://sts.stanford.edu/publications/experience-demand-what-virtual-reality-how-it-works-and-what-it-can-do [accessed 2025-11-03]
2. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Phil Trans R Soc B 2009 Dec 12;364(1535):3549-3557. [doi: 10.1098/rstb.2009.0138]
3. Sanchez-Vives MV, Slater M. From presence to consciousness through virtual reality. Nat Rev Neurosci 2005 Apr;6(4):332-339. [doi: 10.1038/nrn1651]
4. Seth AK, Suzuki K, Critchley HD. An interoceptive predictive coding model of conscious presence. Front Psychology 2012;2. [doi: 10.3389/fpsyg.2011.00395]
5. Kisker J, Gruber T, Schöne B. Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. Psychol Res 2021 Feb;85(1):68-81. [doi: 10.1007/s00426-019-01244-9]
6. Slater M, Wilbur S. A Framework for Immersive Virtual Environments (FIVE): speculations on the role of presence in virtual environments. Presence: Teleoperators & Virtual Environments 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
7. Reichenberger J, Wiggert N, Wilhelm FH, et al. Fear of negative and positive evaluation and reactivity to social-evaluative videos in social anxiety disorder. Behav Res Ther 2019 May;116(140-8):140-148. [doi: 10.1016/j.brat.2019.03.009] [Medline: 30921745]
8. Globisch J, Hamm AO, Esteves F, Ohman A. Fear appears fast: temporal course of startle reflex potentiation in animal fearful subjects. Psychophysiology 1999 Jan;36(1):66-75. [doi: 10.1017/s0048577299970634] [Medline: 10098381]
9. Qi S, Hassabis D, Sun J, Guo F, Daw N, Mobbs D. How cognitive and reactive fear circuits optimize escape decisions in humans. Proc Natl Acad Sci USA 2018 Mar 20;115(12):3186-3191. [doi: 10.1073/pnas.1712314115] [Medline: 29507207]
10. Kuhne C, Kecelioglu ED, Maltby S, et al. Direct comparison of virtual reality and 2D delivery on sense of presence, emotional and physiological outcome measures. Front Virtual Real 2023;4:1211001. [doi: 10.3389/frvir.2023.1211001]
11. Courtney CG, Dawson ME, Schell AM, Iyer A, Parsons TD. Better than the real thing: eliciting fear with moving and static computer-generated stimuli. Int J Psychophysiol 2010 Nov;78(2):107-114. [doi: 10.1016/j.ijpsycho.2010.06.028] [Medline: 20600370]

12. Corr PJ, McNaughton N. Neuroscience and approach/avoidance personality traits: a two stage (valuation-motivation) approach. Neurosci Biobehav Rev 2012 Nov;36(10):2339-2354. [doi: 10.1016/j.neubiorev.2012.09.013] [Medline: 23041073]
13. Daniel-Watanabe L, Fletcher PC. Are fear and anxiety truly distinct? Biol Psychiatry Glob Open Sci 2022 Oct;2(4):341-349. [doi: 10.1016/j.bpsgos.2021.09.006] [Medline: 36324663]
14. Stinson FS, Dawson DA, Patricia Chou S, et al. The epidemiology of DSM-IV specific phobia in the USA: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Psychol Med 2007 Jul;37(7):1047-1059. [doi: 10.1017/S0033291707000086] [Medline: 17335637]
15. Wardenaar KJ, Lim CCW, Al-Hamzawi AO, et al. The cross-national epidemiology of specific phobia in the World Mental Health Surveys. Psychol Med 2017 Jul;47(10):1744-1760. [doi: 10.1017/S0033291717000174] [Medline: 28222820]
16. Hayashi N, Someya N, Maruyama T, Hirooka Y, Endo MY, Fukuba Y. Vascular responses to fear-induced stress in humans. Physiol Behav 2009 Oct 19;98(4):441-446. [doi: 10.1016/j.physbeh.2009.07.008] [Medline: 19619569]
17. Jang EH, Byun S, Park MS, Sohn JH. Reliability of physiological responses induced by basic emotions: a pilot study. J Physiol Anthropol 2019 Nov 28;38(1):15. [doi: 10.1186/s40101-019-0209-y] [Medline: 31779708]
18. Lemmens JS, Simon M, Sumter SR. Fear and loathing in VR: the emotional and physiological effects of immersive games. Virtual Real 2022 Mar;26(1):223-234. [doi: 10.1007/s10055-021-00555-w]
19. Biedermann SV, Biedermann DG, Wenzlaff F, et al. An elevated plus-maze in mixed reality for studying human anxiety-related behavior. BMC Biol 2017 Dec 21;15(1):125. [doi: 10.1186/s12915-017-0463-6] [Medline: 29268740]
20. Zuckerman M. Behavioral Expressions and Biosocial Bases of Sensation Seeking, 1st edition: Cambridge University Press; 1994. URL: https://assets.cambridge.org/97805214/32009/excerpt/9780521432009_excerpt.pdf [accessed 2025-11-03]
21. Solomon Z, Ginzburg K, Neria Y, Ohry A. Coping with war captivity: the role of sensation seeking. Eur J Pers 1995 Mar;9(1):57-70. [doi: 10.1002/per.2410090105]
22. Franken RE, Gibson KJ, Rowland GL. Sensation seeking and the tendency to view the world as threatening. Pers Individ Dif 1992 Jan;13(1):31-38. [doi: 10.1016/0191-8869(92)90214-A]
23. Zuckerman M, Kolin EA, Price L, Zoob I. Development of a sensation-seeking scale. J Consult Psychol 1964;28(6):477-482. [doi: 10.1037/h0040995]
24. Sharifpour M, Walters G, Ritchie BW. The mediating role of sensation seeking on the relationship between risk perceptions and travel behavior. Tourism Analysis 2013 Nov 12;18(5):543-557. [doi: 10.3727/108354213X13782245307795]
25. Lissek S, Powers AS. Sensation seeking and startle modulation by physically threatening images. Biol Psychol 2003 May;63(2):179-197. [doi: 10.1016/s0301-0511(03)00053-x] [Medline: 12738407]
26. Blankstein KR. The sensation seeker and anxiety reactivity: relationships between the sensation-seeking scales and the activity preference questionnaire. J Clin Psychol 1975 Oct;31(4):677-681. [Medline: 1194425]
27. Wagner AM, Houlihan DD. Sensation seeking and trait anxiety in hang-glider pilots and golfers. Pers Individ Dif 1994 Jun;16(6):975-977. [doi: 10.1016/0191-8869(94)90240-2]
28. Monasterio E, Mulder R, Frampton C, Mei-Dan O. Personality characteristics of BASE jumpers. J Appl Sport Psychol 2012;24(4):391-400. [doi: 10.1080/10413200.2012.666710]
29. Poulton R, Davies S, Menzies RG, Langley JD, Silva PA. Evidence for a non-associative model of the acquisition of a fear of heights. Behav Res Ther 1998 May;36(5):537-544. [doi: 10.1016/s0005-7967(97)10037-7] [Medline: 9648329]
30. Zuckerman M, Eysenck S, Eysenck HJ. Sensation seeking in England and America: cross-cultural, age, and sex comparisons. J Consult Clin Psychol 1978 Feb;46(1):139-149. [doi: 10.1037//0022-006x.46.1.139] [Medline: 627648]
31. Arnett J. Sensation seeking: a new conceptualization and a new scale. Pers Individ Dif 1994 Feb;16(2):289-296. [doi: 10.1016/0191-8869(94)90165-1]
32. Nederhof AJ. Methods of coping with social desirability bias: a review. Euro J Social Psych 1985 Jul;15(3):263-280. [doi: 10.1002/ejsp.2420150303]
33. Donaldson SI, Grant-Vallone EJ. Understanding self-report bias in organizational behavior research. J Bus Psychol 2002 Dec;17(2):245-260. [doi: 10.1023/A:1019637632584]
34. Tanzer NK. Bias and equivalence in cross-cultural assessment: an overview. Eur Rev Appl Psychol 1997;47(4):263-280 [FREE Full text]
35. Darkes J, Greenbaum PE, Goldman MS. Sensation seeking-disinhibition and alcohol use: exploring issues of criterion contamination. Psychol Assess 1998;10(1):71-76. [doi: 10.1037//1040-3590.10.1.71]
36. Oberlin BG, Ramer NE, Bates SM, et al. Quantifying behavioral sensation seeking with the aroma choice task. Assessment 2020 Jul;27(5):873-886. [doi: 10.1177/1073191119864659] [Medline: 31353921]
37. Lungwitz EA, Dzemidzic M, Shen YI, Plawecki MH, Oberlin BG. Brain response in heavy drinkers during cross-commodity alcohol and money discounting with potentially real rewards: a preliminary study. Drug Alcohol Depend Rep 2023 Sep;8(100175):100175. [doi: 10.1016/j.dadr.2023.100175] [Medline: 37753349]
38. Oberlin BG, Carron CR, Ramer NE, Plawecki MH, O'Connor SJ, Kareken DA. Intoxication effects on impulsive alcohol choice in heavy drinkers: correlation with sensation seeking and differential effects by commodity. Alcohol Clin Exp Res 2021 Jan;45(1):204-214. [doi: 10.1111/acer.14497] [Medline: 33119917]
39. Camerer C, Mobbs D. Differences in behavior and brain activity during hypothetical and real choices. Trends Cogn Sci (Regul Ed) 2017 Jan;21(1):46-56. [doi: 10.1016/j.tics.2016.11.001]

40. Ebstein RP. The molecular genetic architecture of human personality: beyond self-report questionnaires. Mol Psychiatry 2006 May 1;11(5):427-445. [doi: 10.1038/sj.mp.4001814]
41. Boccignone G, Gadia D, Maggiorini D, Ripamonti LA, Tosto V. Wuthering heights: gauging fear at altitude in virtual reality. Multimed Tools Appl 2023 Feb;82(4):5207-5228. [doi: 10.1007/s11042-022-13366-1]
42. Yilmaz Balban M, Cafaro E, Saue-Fletcher L, et al. Human responses to visually evoked threat. Curr Biol 2021 Feb;31(3):601-612. [doi: 10.1016/j.cub.2020.11.035]
43. Buerke M, Caulfield N, Karnick A, Hill R, Tucker RP, Capron DW. Suicidal ideation and sensation seeking predict suicidal completion in virtual reality: considerations for the future use of virtual reality for the study of suicidality. J technol behav sci 2024;10(1):51-58. [doi: 10.1007/s41347-024-00414-w]
44. Toast_Interactive. Richie's plank experience. 2016. URL: https://toastinteractive.com [accessed 2025-10-30]
45. Speilberger CD, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory: Consulting Psychologists Press; 1983. URL: https://www.researchgate.net/publication/235361542_Manual_for_the_State-Trait_Anxiety_Inventory_Form_Y1_-_Y2 [accessed 2025-11-03]
46. Cohen DC. Comparison of self-report and overt-behavioral procedures for assessing acrophobia. Behav Ther 1977 Jan;8(1):17-23. [doi: 10.1016/S0005-7894(77)80116-0]
47. Aluja A, Rossier J, García LF, Angleitner A, Kuhlman M, Zuckerman M. A cross-cultural shortened form of the ZKPQ (ZKPQ-50-cc) adapted to English, French, German, and Spanish languages. Pers Individ Dif 2006 Sep;41(4):619-628. [doi: 10.1016/j.paid.2006.03.001]
48. Balban MY, Neri E, Kogon MM, et al. Brief structured respiration practices enhance mood and reduce physiological arousal. Cell Rep Med 2023 Jan 17;4(1):100895. [doi: 10.1016/j.xcrm.2022.100895] [Medline: 36630953]
49. Rader N, Bausano M, Richards JE. On the nature of the visual-cliff-avoidance response in human infants. Child Dev 1980 Mar;51(1):61-68. [Medline: 7363749]
50. Gibson EJ, Walk RD. The "Visual Cliff". Sci Am 1960 Apr;202(4):64-71. [doi: 10.1038/scientificamerican0460-64] [Medline: 13827949]
51. Quest 2 tech specs. Meta. URL: https://www.meta.com/quest/products/quest-2/tech-specs [accessed 2025-08-18]
52. Base VEK. How to change the IPD on the Oculus Quest 2?. 2022. URL: https://knowledge.vr-expert.com/kb/how-to-change-the-ipd-on-the-oculus-quest-2/ [accessed 2025-08-18]
53. 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] [Medline: 32300295]
54. Maymon CN, Crawford MT, Blackburne K. The presence of fear: how subjective fear, not physiological arousal, shapes the experience of presence. J Exp Psychol Gen 2024. [doi: 10.31219/osf.io/2dx7n]
55. Kalatzis A, Stanley L, Prabhu VG. Affective state classification in virtual reality environments using electrocardiogram and respiration signals. Presented at: 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR); Nov 15-17, 2021; Taichung, Taiwan. [doi: 10.1109/AIVR52153.2021.00037]
56. Ali HH, Glickman M, Sharot T. The role of habituation in risk-taking escalation. In: Sharot T, editor. Presented at: Conference on Cognitive Computational Neuroscience Proceedings; Aug 24-27, 2023; Oxford, UK URL: https://www.2023.ccneuro.org/proceedings/0001245.pdf?s=W&pn=1633 [accessed 2025-11-03]
57. Caulfield NM, Karnick AT, Capron DW. Exploring dissociation as a facilitator of suicide risk: A translational investigation using virtual reality. J Affect Disord 2022 Jan 15;297:517-524. [doi: 10.1016/j.jad.2021.10.097] [Medline: 34715163]
58. Smith CS, Warner M. Cyberdelics: context engineering psychedelics for altered traits. Presented at: Proceedings of EVA; Jul 4-8, 2022; London, UK URL: https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/EVA2022.48 [accessed 2025-11-03]
59. Ball IL, Farnill D, Wangeman JF. Sex and age differences in sensation seeking: some national comparisons. British J of Psychology 1984 May;75(2):257-265. [doi: 10.1111/j.2044-8295.1984.tb01897.x]
60. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995 Jan 1;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
61. Rothbaum BO, Hodges LF, Ready D, Graap K, Alarcon RD. Virtual reality exposure therapy for Vietnam veterans with posttraumatic stress disorder. J Clin Psychiatry 2001 Aug;62(8):617-622. [doi: 10.4088/jcp.v62n0808] [Medline: 11561934]
62. Lee JH, Ku J, Kim K, et al. Experimental application of virtual reality for nicotine craving through cue exposure. Cyberpsychol Behav 2003 Jun;6(3):275-280. [doi: 10.1089/109493103322011560]
63. van Loenen I, Scholten W, Muntingh A, Smit J, Batelaan N. The effectiveness of virtual reality exposure-based cognitive behavioral therapy for severe anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder: meta-analysis. J Med Internet Res 2022 Feb 10;24(2):e26736. [doi: 10.2196/26736] [Medline: 35142632]
64. Hoffman HG, Patterson DR, Carrougher GJ, Sharar SR. Effectiveness of virtual reality-based pain control with multiple treatments. Clin J Pain 2001 Sep;17(3):229-235. [doi: 10.1097/00002508-200109000-00007] [Medline: 11587113]
65. Hoffman HG, Garcia-Palacios A, Kapa V, Beecher J, Sharar SR. Immersive virtual reality for reducing experimental ischemic pain. Int J Hum Comput Interact 2003 Jun;15(3):469-486. [doi: 10.1207/S15327590IJHC1503_10]

66. Huang Q, Lin J, Han R, Peng C, Huang A. Using virtual reality exposure therapy in pain management: a systematic review and meta-analysis of randomized controlled trials. Value Health 2022 Feb;25(2):288-301. [doi: 10.1016/j.jval.2021.04.1285]
67. van Gelder JL, Cornet LJM, Zwalua NP, Mertens ECA, van der Schalk J. Interaction with the future self in virtual reality reduces self-defeating behavior in a sample of convicted offenders. Sci Rep 2022 Feb 10;12(1):2254. [doi: 10.1038/s41598-022-06305-5] [Medline: 35145177]
68. Shen YI, Nelson AJ, Oberlin BG. Virtual reality intervention effects on future self-continuity and delayed reward preference in substance use disorder recovery: pilot study results. Discov Ment Health 2022;2(1):19. [doi: 10.1007/s44192-022-00022-1] [Medline: 36128578]
69. Schröder D, Wrona KJ, Müller F, Heinemann S, Fischer F, Dockweiler C. Impact of virtual reality applications in the treatment of anxiety disorders: a systematic review and meta-analysis of randomized-controlled trials. J Behav Ther Exp Psychiatry 2023 Dec;81:101893. [doi: 10.1016/j.jbtep.2023.101893] [Medline: 37453405]
70. Wray TB, Emery NN. Feasibility, appropriateness, and willingness to use virtual reality as an adjunct to counseling among addictions counselors. Subst Use Misuse 2022 Jul 29;57(9):1470-1477. [doi: 10.1080/10826084.2022.2092148]
71. Arissen C, van der Helm L, Dijkstra B, Markus W. Virtual reality to support inpatient addiction treatment: patients are ready, what about therapists? A feasibility study. Int J Ment Health Addiction 2023 Dec;21(6):4085-4107. [doi: 10.1007/s11469-022-00843-9]
72. Childs E, Astur RS, de Wit H. Virtual reality conditioned place preference using monetary reward. Behav Brain Res 2017 Mar 30;322(Pt A):110-114. [doi: 10.1016/j.bbr.2017.01.019] [Medline: 28108321]
73. Elliot AJ. The hierarchical model of approach-avoidance motivation. Motiv Emot 2006 Oct 6;30(2):111-116. [doi: 10.1007/s11031-006-9028-7]
74. Bracha HS, Bienvenu OJ, Eaton WW. Testing the Paleolithic-human-warfare hypothesis of blood-injection phobia in the Baltimore ECA Follow-up Study—towards a more etiologically-based conceptualization for DSM-V. J Affect Disord 2007 Jan;97(1-3):1-4. [doi: 10.1016/j.jad.2006.06.014] [Medline: 16860872]
75. MacDonald DH, Hewlett BS. Reproductive interests and forager mobility. Curr Anthropol 1999 Aug;40(4):501-524. [doi: 10.1086/200047]
76. Miner EJ, Gurven M, Kaplan H, Gaulin SJC. Sex difference in travel is concentrated in adolescence and tracks reproductive interests. Proc Biol Sci 2014 Dec 7;281(1796):20141476. [doi: 10.1098/rspb.2014.1476] [Medline: 25320169]
77. Sear R, Mace R. Who keeps children alive? A review of the effects of kin on child survival. Evol Hum Behav 2008 Jan;29(1):1-18. [doi: 10.1016/j.evolhumbehav.2007.10.001]
78. McKibbin WF, Shackelford TK. Women's avoidance of rape. Aggress Violent Behav 2011 Sep;16(5):437-443. [doi: 10.1016/j.avb.2011.03.010]
79. Loijen A, Vrijsen JN, Egger JIM, Becker ES, Rinck M. Biased approach-avoidance tendencies in psychopathology: a systematic review of their assessment and modification. Clin Psychol Rev 2020 Apr;77:101825. [doi: 10.1016/j.cpr.2020.101825] [Medline: 32143108]
80. Volkow ND, Wang GJ, Ma Y, et al. Activation of orbital and medial prefrontal cortex by methylphenidate in cocaine-addicted subjects but not in controls: relevance to addiction. J Neurosci 2005 Apr 13;25(15):3932-3939. [doi: 10.1523/JNEUROSCI.0433-05.2005] [Medline: 15829645]
81. Mujica-Parodi LR, Carlson JM, Cha 차지욱 J, Rubin D. The fine line between "brave" and "reckless": amygdala reactivity and regulation predict recognition of risk. NeuroImage 2014 Dec;103:1-9. [doi: 10.1016/j.neuroimage.2014.08.038] [Medline: 25175540]
82. Etkin A, Wager TD. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. Am J Psychiatry 2007 Oct;164(10):1476-1488. [doi: 10.1176/appi.ajp.2007.07030504] [Medline: 17898336]
83. Elman I, Lowen S, Frederick BB, Chi W, Becerra L, Pitman RK. Functional neuroimaging of reward circuitry responsivity to monetary gains and losses in posttraumatic stress disorder. Biol Psychiatry 2009 Dec 15;66(12):1083-1090. [doi: 10.1016/j.biopsych.2009.06.006] [Medline: 19640506]
84. Zuckerman M, Neeb M. Demographic influences in sensation seeking and expressions of sensation seeking in religion, smoking and driving habits. Pers Individ Dif 1980;1(3):197-206. [doi: 10.1016/0191-8869(80)90051-3]
85. Wang W, Wu YX, Peng ZG, et al. Test of sensation seeking in a Chinese sample. Pers Individ Dif 2000 Jan;28(1):169-179. [doi: 10.1016/S0191-8869(99)00092-6]
86. Segal B. Sensation seeking and anxiety: assessment of responses to specific stimulus situations. J Consult Clin Psychol 1973 Aug;41(1):135-138. [doi: 10.1037/h0035594] [Medline: 4726696]
87. Castro-Toledo FJ, Perea-García JO, Bautista-Ortuño R, Mitkidis P. Influence of environmental variables on fear of crime: comparing self-report data with physiological measures in an experimental design. J Exp Criminol 2017 Dec;13(4):537-545. [doi: 10.1007/s11292-017-9295-1]
88. McLeod DR, Hoehn-Saric R, Stefan RL. Somatic symptoms of anxiety: comparison of self-report and physiological measures. Biol Psychiatry 1986 Mar;21(3):301-310. [doi: 10.1016/0006-3223(86)90051-x] [Medline: 3947711]
89. Edelmann RJ, Baker SR. Self‐reported and actual physiological responses in social phobia. British J Clinic Psychol 2002 Mar;41(1):1-14. [doi: 10.1348/014466502163732]

90. Garfinkel SN, Seth AK, Barrett AB, Suzuki K, Critchley HD. Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. Biol Psychol 2015 Jan;104:65-74. [doi: 10.1016/j.biopsycho.2014.11.004]
91. Cyders MA, Coskunpinar A. Measurement of constructs using self-report and behavioral lab tasks: is there overlap in nomothetic span and construct representation for impulsivity? Clin Psychol Rev 2011 Aug;31(6):965-982. [doi: 10.1016/j.cpr.2011.06.001] [Medline: 21733491]
92. Murphy BA, Lilienfeld SO. Are self-report cognitive empathy ratings valid proxies for cognitive empathy ability? Negligible meta-analytic relations with behavioral task performance. Psychol Assess 2019 Aug;31(8):1062-1072. [doi: 10.1037/pas0000732] [Medline: 31120296]
93. Frey R, Pedroni A, Mata R, Rieskamp J, Hertwig R. Risk preference shares the psychometric structure of major psychological traits. Sci Adv 2017 Oct;3(10):e1701381. [doi: 10.1126/sciadv.1701381] [Medline: 28983511]
94. Messina A, Berntsen D. Self-reported sensibility to bodily signals predicts individual differences in autobiographical memory: an exploratory study. Memory 2024 Sep;32(8):996-1011. [doi: 10.1080/09658211.2024.2373891] [Medline: 38990765]
95. Lang B. A comparison of risk-taking measures. J Undergrad Res Minn State Univ Mankato 2011;11(1):3. [doi: 10.56816/2378-6949.1025]
96. Dang J, King KM, Inzlicht M. Why are self-report and behavioral measures weakly correlated? Trends Cogn Sci (Regul Ed) 2020 Apr;24(4):267-269. [doi: 10.1016/j.tics.2020.01.007]
97. Dunning D, Heath C, Suls JM. Flawed self-assessment: implications for health, education, and the workplace. Psychol Sci Public Interest 2004 Dec;5(3):69-106. [doi: 10.1111/j.1529-1006.2004.00018.x] [Medline: 26158995]
98. Kelley TL. Interpretation of Educational Measurements: World Book Company; 1927. URL: https://gwern.net/doc/iq/1927-kelley-interpretationofeducationalmeasurements.pdf [accessed 2025-11-03]

Abbreviations

ACT: Aroma Choice Task
STAI-S: State Trait Anxiety Inventory-State
VR: virtual reality

Edited by T Leung, L Riedemann; submitted 12.03.25; peer-reviewed by BS Khundrakpam, HJ Suk, V Vrushali; revised version received 01.10.25; accepted 23.10.25; published 04.11.25.

Please cite as:
Daugherty RC, Jacksack CS, Daye BJ, Oberlin BG
Association Between Sensation Seeking and Fear Response: Interventional Study of Personality and Behavior Using a Virtual Reality Heights Simulation
JMIR XR Spatial Comput 2025;2:e73785
URL: https://xr.jmir.org/2025/1/e73785
doi:10.2196/73785

Extended Reality Biofeedback for Functional Upper Limb Weakness: Mixed Methods Usability Evaluation

Anirban Dutta[1], PhD; Katerina Hatjipanagioti, BA (Hons); Matthew Alexander Newsham[2]; Lewis Leyland[3]; Lindsey Rickson[3], BSc (Hons), HCPC, MCSP; Alastair Buchanan[4], PhD; Ildar Farkhatdinov[5], PhD; Jacqueline Twamley[3], RGN, BSc (Hons), MSc, PhD; Abhijit Das[6,7], MD, DM

[1]Department of Metabolism and Systems Science, University of Birmingham, Edgbaston, Birmingham, United Kingdom

[2]FND North, Heysham Morecambe, United Kingdom

[3]Royal Preston Hospital, Preston, United Kingdom

[4]Nudge Reality Ltd, Cheshire, United Kingdom

[5]School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom

[6]Lancashire Teaching Hospitals NHS Foundation Trust, Preston, United Kingdom

[7]University of Central Lancashire, Preston, United Kingdom

Corresponding Author:
Anirban Dutta, PhD
Department of Metabolism and Systems Science, University of Birmingham, Edgbaston, Birmingham, United Kingdom

Abstract

Background: The perception–action cycle enables humans to adapt their behaviors by integrating sensory feedback into motor actions. Functional neurological disorder (FND) disrupts this cycle, leading to maladaptive motor responses and a diminished sense of agency. FND includes functional seizures, movement disorders, and cognitive impairments, significantly affecting quality of life. Recent advancements in extended reality (XR) neurotechnologies provide opportunities for novel rehabilitation approaches, leveraging visual and haptic feedback to retrain motor control and restore agency in individuals with functional limb weakness.

Objective: This study aimed to co-design and evaluate an XR-based biofeedback platform for upper-limb rehabilitation in FND, incorporating multisensory feedback (visual and haptic) to enhance motor retraining.

Methods: A mixed methods design was used. In phase 1, a Delphi survey (N=20, patients with FND) identified key user requirements, emphasizing customizability, real-time feedback, accessibility, and comfort. These insights guided the codevelopment of an XR biofeedback platform. In phase 2, a co-design workshop with 6 participants (3 FND patient representatives and 3 health care professionals) evaluated the usability of 3 XR training tasks: virtual reality (VR) relaxation task, a guided meditation in a VR calming environment; XR position feedback task (“Hoop Hustle”), a VR-based motion task requiring arm movements to interact with virtual objects, providing real-time positional biofeedback; and XR force feedback task, a haptic robot-assisted exercise using the Human Robotix System (HRX-1) haptic device, applying resistive forces to guide upper limb movements. Participants completed system usability scale (SUS) questionnaires and provided qualitative feedback, which was analyzed using NVivo (QSR International) thematic analysis.

Results: The XR position feedback task achieved the highest usability ratings, with 4 out of 6 participants scoring it above 85, indicating “excellent” usability. The VR relaxation task received polarized scores: 2 participants rated it highly (90 and 87.5), while 3 scored it poorly (mid-40s), citing motion discomfort and disengagement. The XR force feedback task had mixed usability outcomes (SUS range: 27.5 - 95.0), with 1 participant with functional dystonia struggling significantly (SUS 27.5), while others rated it between 62.5 and 95.0. Qualitative feedback emphasized comfort (lighter headsets and better ergonomic design), immersion and content quality (clearer visuals and reduced distracting audio prompts), personalization (adjustable settings for speed, difficulty, and force resistance), and accessibility (cost concerns and home usability considerations). Overall, participants viewed the XR biofeedback platform as highly promising but in need of fine-tuning.

Conclusions: This study demonstrates the feasibility and usability of an XR neurotechnology platform for FND rehabilitation, with strong acceptance of XR position feedback, mixed reactions to VR relaxation, and individual-specific usability outcomes for the force feedback task. Findings underscore the need for personalization features and hardware refinement. Future work will focus on enhancing usability, improving accessibility, and evaluating effectiveness in larger clinical trials.

(JMIR XR Spatial Comput 2025;2:e68580) doi:10.2196/68580

KEYWORDS

extended reality; haptics; functional neurological disorder; biofeedback; usability; co-design; System Usability Scale

Introduction

Humans continuously learn through interactions with their environment via a perception-action cycle—a feedback loop where sensory input informs actions and the consequences of these actions (shaped by rewards and penalties) reinforce or modify behavior over time. This adaptive learning process is crucial for navigating social and environmental contexts, allowing individuals to align their behaviors with societal norms and expectations. However, maladaptive learning can occur when responses to rewards and penalties lead to dysfunctional behavior patterns, diminishing an individual's sense of agency and resulting in disordered actions [1]. We hypothesize that functional neurological disorder (FND) may arise from such maladaptive learning within the perception-action cycle, where certain reinforced behaviors disrupt normal functional responses, contributing to symptoms and reduced voluntary control over bodily actions.

FND is a complex, debilitating condition with symptoms comparable in severity and societal cost to those of epilepsy or multiple sclerosis [2]. FND encompasses several subtypes—functional seizures, functional movement disorders, persistent perceptual postural dizziness, and functional cognitive disorder—stemming from interplay between neurological and psychological mechanisms [3]. Yet, only about 50% of United Kingdom health boards have established care pathways for FND, underscoring significant gaps in treatment [4]. Recent advancements in neurotechnology and better understanding of FND pathophysiology have revealed shared mechanisms (such as abnormal sensorimotor processing and disruptions in sense of agency) that can be targeted by novel therapeutic strategies [3]. Notably, extended reality (XR) approaches have been proposed within a stepped-care rehabilitation framework [5], enabling interventions to be tailored based on symptom severity and delivered from clinic to home settings. XR is an umbrella term encompassing immersive technologies that blend digital and physical environments, including augmented reality (AR), virtual reality (VR), and mixed reality (MR). AR overlays digital information onto the real world, VR fully immerses users in a computer-generated environment, and MR allows interactive overlay of artificial elements onto the real world. XR platforms can incorporate haptic (touch-based) feedback and guided suggestions to engage patients through bottom-up sensory input and top-down cognitive cues, respectively, aiming to retrain the disrupted perception-action links underlying FND symptoms [6]. For example, haptic feedback may provide real-time physical cues to encourage movement, while positive verbal reinforcement ("You're doing great!") can facilitate operant conditioning during VR rehabilitation [7].

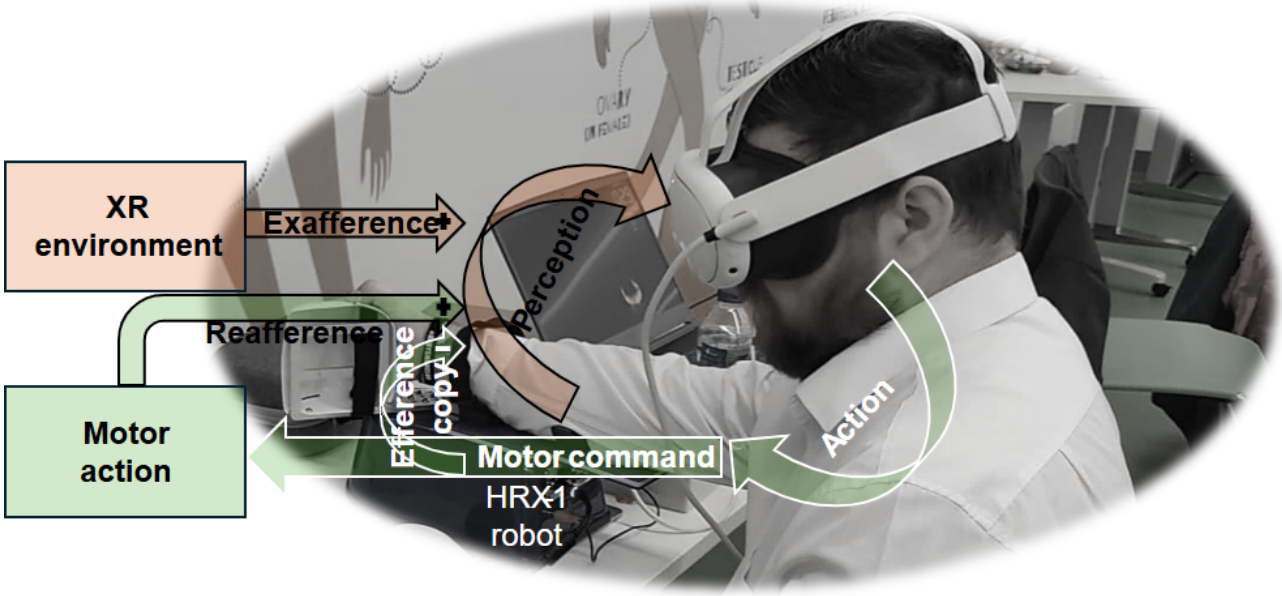
A survey of 527 individuals revealed high comorbidity rates among patients with FND, with pain (78.1%), fatigue (78.0%), and sleep disturbances (46.7%) being the most common symptoms, often worsening postdiagnosis [8]. Effective FND management underscores the need for transparent diagnosis explanations to improve patient understanding and enable personalized treatment strategies [9]. The National Institute of Mental Health's Research Domain Criteria framework [10] offers a dimensional perspective for understanding FND [11], guiding the development of neurotechnologies and biomarkers to better categorize its heterogeneity. The recent proposal for the inclusion of the sensorimotor domain in the Research Domain Criteria highlights the growing recognition of sensorimotor processing in mental health [10], presenting opportunities for intervention through XR neurotechnologies.

Building on previous VR-based interventions [12], we proposed the integration of haptic feedback into an XR setting to modulate the balance between sensory attenuation and amplification using an operant conditioning framework [7]. Haptic feedback in visuo-motor tasks plays a crucial role in reinforcing the perception-action cycle, primarily through efference copy and corollary discharge integration, which differs from motor imagery-based VR training [13]. The efference copy is an internal duplicate of motor commands from the supplementary motor complex [14], allowing the cerebellum and sensory areas to predict sensory consequences of movement [15]. This predictive function enables the brain to distinguish between self-generated actions and external stimuli, an essential aspect of sensorimotor learning. When haptic feedback is absent, motor learning relies on mental simulations without new sensory data, potentially reinforcing maladaptive internal models, as observed in cerebellar dysfunction [13,16]. In adaptive XR learning, haptic feedback serves as real-world sensory input, aiding in the recalibration of maladaptive internal models and reducing overreliance on predictive mechanisms associated with mental simulations in VR-only settings. Studies show that without haptic input, individuals struggle to correct motor prediction errors, as their internal model fails to recalibrate effectively [17]. By integrating haptic feedback into XR rehabilitation, we aim to recalibrate maladaptive sensorimotor patterns related to fatigue (effort-reward mismatch [18]), pain, weakness, dystonia, and seizures.

Support for XR-based functional motor disorder (FMD) rehabilitation also stems from intentional binding research, which suggests that repeated operant experiences enhance implicit agency by reinforcing associative learning [19]. This highlights the distinction between explicit and implicit agency: explicit agency, tied to conscious awareness, can be strengthened through demonstrations like Hoover's sign or tremor entrainment [20], while implicit agency is shaped through repeated operant conditioning [7]. These mechanisms interact via top-down and bottom-up pathways, which can be experimentally modulated in XR through exafference—the controlled simulation of external stimuli. However, the ethical, cost, and usability concerns associated with digital health interventions necessitate stakeholder engagement to ensure alignment with broader health care goals. Industry-driven digital health innovation plays a key role in assessing how these technologies impact health care systems and patient outcomes. Our research focuses on evaluating the usability of an XR neurotechnology platform for biofeedback training in functional limb weakness, combining

bottom-up haptic feedback with top-down visuo-motor task suggestions (refer to Figure 1) [6]. The ultimate goal is to develop precise, technically effective, sustainable, and patient-friendly XR neurotechnologies for FND rehabilitation. Industry-driven innovation plays a key role in translating these technologies into practice by evaluating their impact on health care systems and outcomes. The ultimate objective is to ensure that such neurotechnology is not only effective but also user-friendly, acceptable, and accessible for people with FND. In this context, this study adopted a coproduction approach to co-design an XR biofeedback training platform for functional limb weakness in FND and to assess its usability with end-users.

Figure 1. Perception-action coupling for extended reality (XR) biofeedback training to modulate bottom-up reafference with exafference through a haptic robot (HRX-1) to support movement in cases of functional weakness. Top-down modulation is influenced by guided visual and verbal suggestions presented via XR feedback. A distinction can be made between efference copy—internal brain duplicates of motor commands (in action)—and corollary discharge, which involves expected sensory signals due to those motor commands (in perception).



Methods

Study Setting and Participants

This study consisted of two phases: an exploratory survey (Delphi method) conducted online to inform platform design and a subsequent in-person co-design workshop for usability evaluation.

In phase 1, an exploratory Delphi survey was conducted online, where a convenience sampling method was used to recruit individuals with lived experience of FND as "experts by experience" from the United Kingdom Royal Preston Hospital's FND service team's networks led by the PPIE (Patient and Public Involvement and Engagement) leads. In total, 20 individuals (experts by experience) with FND participated in the initial round of the Delphi survey. Participants provided feedback via an online questionnaire. The survey collected both quantitative and qualitative data on several topics: familiarity with VR and haptic technologies, perceptions of comfort and ease of use, anticipated relevance and impact of an XR-based therapy for FND, and potential barriers to adoption (such as, cost, access to equipment, technical support, and side effects). Responses were analyzed to extract common themes and requirements that the PPIE lead presented at the National Rehabilitation Centre (NRC) Rehabilitation Technologies Conference 2024 [21] (NRC Rehabilitation Technologies Conference 2024 poster and slides in Multimedia Appendix 1). Based on the survey findings, we codeveloped with the PPIE leads and industry partners (Human Robotix Ltd and Nudge Reality Ltd) a prototype XR neurotechnology platform. We selected the Human Robotix HRX-1 upper-limb haptic system (a portable robotic device providing force feedback) and Nudge Reality's "Hoop Hustle" XR game as the core components for our platform, as these were judged by the PPIE leads to best meet the identified needs (detailed specifications of the hardware and game options are available in Multimedia Appendix 2). The HRX-1 device can assist or resist arm movements with precise torque control, while Hoop Hustle is a VR game that can be adapted for therapeutic exercises.

For the phase 2 co-design and usability testing workshop, a purposive sampling approach was used to recruit participants specifically from the United Kingdom Royal Preston Hospital's FND Service team, including PPIE leads. We then conducted an in-person workshop involving 6 participants drawn from the FND service community: 3 FND patient representatives (1 female and 2 male) and 3 health care professionals (2 physiotherapists and 1 neurologist; 2 female and 1 male). All 6 participants are coauthors of this paper for the participatory design approach. Before the workshop, participants provided informed consent. The session took place in a rehabilitation clinic setting and lasted about half a day.

Ethical Considerations

As this work was part of a patient engagement and technology co-design project, it was conducted with institutional review board notification but was determined to be a service development and quality improvement activity not requiring full National Health Service (NHS) Research Ethics Committee

review. All participants gave written informed consent for their involvement and for publication of deidentified feedback. The study was carried out in accordance with the Declaration of Helsinki principles of ethical research.

Procedure and XR Platform Tasks

Given the selection of Human Robotix's HRX-1 system for upper limb rehabilitation (Human Robotix's HRX-1 system in Multimedia Appendix 3) and Nudge Reality's "Hoop Hustle" game (Nudge Reality's XR games in Multimedia Appendix 4) by PPIE leads, efforts were focused on adapting these technologies to test 3 conditions: VR relaxation, XR positional feedback, and XR force feedback. During the co-design workshop, the prototype XR platform was introduced, and participants were guided through 3 interactive training tasks, each representing a different mode of biofeedback.

Experimental Robotic System

A 1-degree-of-freedom HRX-1 desktop robot (refer to Figure 2) equipped with a direct-drive electromagnetic motor for wrist flexion or extension movement was used in the study. The robot offers high flexion or extension torque (up to 2 Nm), position and torque sensing, and a variety of control modes in a compact robotic platform. The design of the HRX-1 robot is substantially more compact and lighter than existing comparable systems to enable easy transportation and installation for the studies in clinical, research, and at-home environments. The safety of the robot operation was implemented at mechanical (range of motion limitation with end-stops), electric (limitation for the maximal electric current), and software (limitation on the maximal speed of movement) levels. Previously, robots have been successfully used in clinical and research studies [22-24]. In this study, the HRX-1 robot was integrated with VR tasks.

Figure 2. HRX-1 robot that can generate programmable wrist flexion and extension torques for assistance or resistance during the experimental study.

VR Relaxation Task

Participants wore a Meta Quest 3 VR headset to experience a guided relaxation session. The VR environment featured calming scenery (eg, a gradually descending landscape or serene nature scene), accompanied by a gentle narrative instructing the user in relaxation techniques (for instance, breathing exercises, and progressive muscle relaxation cues). The purpose of this task was to familiarize users with VR and induce relaxation, which can help reduce FND symptom intensity. Participants remained seated during this task. Notably, based on user feedback from the Delphi survey, we avoided any instructions that would conflict with VR immersion (as one Delphi respondent cautioned that this could cause disorientation). The task lasted about 5 - 7 minutes.

XR Position Feedback Game (Hoop Hustle)

In this task, participants engaged with hoop hustle, a therapeutic game developed for XR rehabilitation. The user's goal in the game is to move their affected arm (or a controller held in that arm) to "shoot" balls in VR through a series of hoops or targets at varying positions. The game provides real-time visual

feedback on the accuracy and speed of the user's arm movements. For example, when a participant moves their arm, a corresponding arm or cursor in VR is shown, allowing them to adjust their movement to align with the hoop. Successful hits (getting the ball through the hoop) trigger immediate positive feedback (visual effects and encouraging sounds). The game's difficulty can be adjusted—for example, hoop height and size can be modified to accommodate the user's range of motion, and the speed of ball generation can be tuned. During the workshop, an operator adjusted these settings as needed to ensure each participant could comfortably attempt the task. This task emphasized positional biofeedback (augmented visual feedback of movement) without additional force resistance. Each participant practiced for several minutes until they felt they had experienced the core mechanics of the game.

XR Force Feedback Task

The HRX-1 haptic robot was integrated with the hoop hustle game to provide force feedback during the exercise. Participants grasped the end-effector of the HRX-1 device, which was programmed to apply gentle resistive forces or assistance during specific arm movements in the VR game. For instance, as a participant guided a ball toward a hoop in VR, the device might add a slight downward resistance, requiring the user to exert additional effort and thus engage proprioceptive feedback pathways. In this way, the XR force feedback task combined visual and haptic biofeedback. We also implemented a simple exercise game: the wrist handle of the robot was used to control a visual cursor shown in the screen, and a participant's task was to rotate the handle with their wrist follow a pseudo-random movement of a target on the screen as accurate and as fast as possible, similar to the tasks used in [25]. A participant could observe the progress task on the screen (visual modality) and feel the assistive and resistive wrist flexion or extension torques generated by the robot (force feedback modality). This was included to explore how force feedback might help reveal or train aspects of motor control in FND (eg, addressing sensory attenuation deficits). Each participant spent around 5 minutes with force feedback enabled. One participant with functional dystonia required a brief rest during this task due to muscle fatigue; however, all participants were able to attempt the task to some extent.

Throughout the session, participants were encouraged to "think aloud" and share any difficulties or observations (eg, if the headset felt uncomfortable or if a task was confusing). A facilitator took notes on these observations to supplement the formal feedback.

Data Collection and Analysis

After completing all 3 tasks, participants filled out the system usability scale (SUS) questionnaire for each task. The SUS is a 10-item questionnaire yielding a score from 0 to 100, where higher scores indicate better perceived usability. We chose the SUS because it is a well-established, quick tool for usability assessment, suitable even for small samples [26]. Participants also provided written free-text feedback on their experience with each task and the overall platform. These responses were collected on paper forms and later transcribed. In addition, the workshop concluded with a short group discussion, allowing participants to collectively reflect on what aspects of the platform worked well and what improvements they would prioritize. The discussion was later summarized in notes.

Quantitative data from the SUS were summarized using descriptive statistics, given the small sample size. We report individual SUS scores per participant and per task, as well as the range and median for each task's scores. Following convention [27], we interpret SUS scores using an adjective rating scale for context: scores above~85 are considered "excellent," around 70 - 85 "good,"~50 - 69 "okay," and below 50 "poor" in terms of usability perception. We did not perform inferential statistical tests due to the exploratory nature of this pilot and the limited number of subjects. Qualitative data (written feedback and facilitator notes) were analyzed thematically. Two researchers (1 patient representative and 1 study investigator) independently reviewed the feedback to identify recurring themes. Using NVivo 12 (QSR International), feedback comments were coded with initial labels corresponding to aspects of user experience (eg, "hardware discomfort," "audio feedback," and "game difficulty"). These codes were then grouped into higher-level themes through discussion and consensus. Representative participant quotes were extracted to illustrate each theme in the Results.

Results

Phase 1: Exploratory Delphi Survey Report (CHERRIES Checklist)

We present the results from our first round of the Delphi survey according to the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [28], aimed at the translation of the VR haptics technology for biofeedback training in FND. The survey gathered online feedback from 20 (N=20) individuals with lived experience of FND, considered experts by experience for technology translation.

Design

This was an online Delphi survey aimed at gathering high-level user requirements for the development of a VR haptics biofeedback training platform for FND rehabilitation.

The survey sought to assess perceptions and expectations of VR and haptic biofeedback technology for rehabilitation, potential benefits and usability considerations for upper and lower limb motor retraining, and barriers to adoption and accessibility concerns among individuals with lived experience of FND.

Development and Pretesting

Survey Development

The survey was co-designed by a multidisciplinary team (co-authors of this report), including clinicians, researchers, industry partners, and FND patient representatives. It was pilot-tested with a small group of patients with FND and clinicians to refine clarity, content, and usability.

Survey Refinements

Feedback from pilot testing led to revisions in question phrasing, response categories, and survey logic. Adjustments were made

to ensure accessibility for individuals with neurological impairments (eg, clear navigation and avoiding long response formats).

Recruitment Process and Sample Characteristics

Target Population

The survey targeted adults (≥18 y) with FND, particularly those experiencing functional limb weakness.

Recruitment Strategy

Participants were recruited through FND patient advocacy organizations (eg, FND Hope, FND Action). Neurology clinics specializing in FND care. Online FND support groups and social media communities. The survey link was shared via email, social media, and organizational websites.

Participation Details

Participation details included a survey link access, in which an open-access URL was provided with IP duplicate detection enabled. No monetary incentives were provided; participants were thanked for their contributions in follow-up communications.

Survey Administration

The survey was hosted on a General Data Protection Regulation (GDPR)-compliant, secure online platform (MS Forms is part of Microsoft 365, which adheres to GDPR, Health Insurance Portability and Accountability Act (HIPAA), and ISO 27001 security standards).

Response Tracking

Anonymous participation was allowed; no email registration was required, and no IP tracking or cookies were used.

Survey Content

The question structure of the survey included a combination of question types, as listed in Textbox 1.

Textbox 1. Question structure of the survey.

- Demographics (age, gender, FND diagnosis history, and previous XR or VR experience).
- Experience with VR or haptic technology (previous use in gaming, therapy, etc).
- Perceived benefits of XR biofeedback (customizability, real-time feedback, and usability).
- Barriers to adoption (cost, accessibility, and concerns about motion sickness).
- Open-ended qualitative feedback (expectations, concerns, and usability considerations).

Data Handling and Statistical Analysis

Data Privacy Measures

No personally identifiable information was collected. Responses were stored in a secure, encrypted database, accessible only to authorized researchers.

Analysis Methods

Descriptive statistics were used for Likert-scale responses (percentages and means). Qualitative thematic analysis was performed using NVivo for open-ended responses.

Results Reporting

Response Rate

Response rates are described in Textbox 2.

Textbox 2. Response rate.

- Total respondents: 20.
- Completion rate: 85% (17 fully completed responses).
- Dropout rate: 15% (3 partial responses).

Key Findings

Key findings are mentioned in Textbox 3.

Textbox 3. Key findings.

- Participant demographics:
  - Peak age group: 35-44 years.
- Gender: predominantly female.
- Experience and perception of VR and haptic technology
  - Awareness of VR technology: high, but varied levels of familiarity.
  - Haptic technology experience: less common.
- Comfort levels: mostly positive, but some concerns about mask and goggle discomfort and motion sickness.
- Perceived relevance and potential impact: high perceived relevance for FND rehabilitation.
- Participants prioritized:
  - Customizable exercises.
  - Real-time biofeedback.
  - Immersive environments.
- Barriers and challenges identified
  - Accessibility concerns: (1) cost of VR equipment, (2) availability through NHS or insurance coverage, (3) WiFi or connectivity limitations.
- Usability issues:
  - Motion sickness concerns.
  - Need for guidance on using XR biofeedback at home.
- Potential safety concerns:
  - Risk of falls or overstimulation.

Discussion of Bias and Limitations

Potential Biases

The two types of potential biases are (1) selection bias: participants were self-selected, possibly favoring tech-savvy individuals, or those already engaged in FND support groups; and (2) response bias: some participants may have been overly optimistic or cautious in their feedback.

Limitations

This study has two limitations. The first is the small sample size (N=20); the results are preliminary and not generalizable to all patients with FND. The second is the use of the single-round Delphi survey; the findings require further validation through additional rounds or larger-scale studies.

Conclusion

The first round of the Delphi survey provided key insights into the usability, expectations, and barriers associated with XR haptics biofeedback training for FND rehabilitation.

Key Takeaways

Participants perceived high potential benefits but highlighted cost, accessibility, and usability concerns. There was a strong interest in real-time feedback and customization to tailor the technology to individual needs. Concerns about motion sickness, equipment comfort, and NHS availability need to be addressed for successful adoption.

Future Steps

Refining usability features based on patient feedback in phase 2 co-design and usability testing. Further stakeholder engagement with clinicians, patient organizations, and industry partners in phase 2 co-design and usability testing. Scaling the study to validate findings with a larger sample and additional Delphi rounds following in phase 2 co-design and usability testing.

Phase 2: Usability Scores (Quantitative Results)

Basic usability testing typically benefits from the purposive selection of 5‐10 participants [29]. Here, all 6 workshop participants completed the XR position feedback and XR force feedback tasks, and 5 completed the VR relaxation task (1 health care professional was unable to try the VR relaxation due to time constraints). Table 1 presents the SUS scores given by each participant for each task. Overall, the XR position feedback game received the highest ratings with a median score of 91.3, and all participants rated it above 70. The VR relaxation task had a bimodal distribution of scores—2 participants rated it very highly (~88‐90) while 3 participants gave it scores below 50, indicating poor usability for those individuals. The XR force feedback task had generally positive scores from 4 participants

(range, 80.0 - 95.0), but 1 participant (Participant 3) gave a very low score (27.5). According to Bangor et al's [27] adjective rating scale for SUS, the low scores in the 40s for the VR relaxation task correspond to a "poor" usability experience, despite the same task being rated as "excellent" by others. Similarly, the force feedback task's scores suggest mostly "good" to "excellent" usability, with one clear outlier in the "poor" range. In contrast, the XR position feedback task's scores correspond to "good" or "excellent" usability across all users. These results highlight a high degree of variability in user experience for the more complex or condition-sensitive tasks (VR relaxation and force feedback), compared to the consistently positive experience with the position feedback game (XR system usability testing script and XR system usability testing results in Multimedia Appendices 5-8).

Table . The system usability scale (SUS) score was calculated for each participant across extended reality (XR) tasks, including virtual reality (VR) relaxation, XR position feedback control, and XR force feedback control. Participant 6 did not participate in the VR relaxation task.

	XR force feedback SUS	XR position feedback SUS	VR relaxation SUS
Participant 1	80.0	95.0	47.5
Participant 2	62.5	72.5	90.0
Participant 3	27.5	92.5	45.0
Participant 4	82.5	100.0	87.5
Participant 5	95.0	90.0	45.0
Participant 6	87.5	85.0	—[a]

[a]not available.

Phase 2: User Feedback and Thematic Analysis (Qualitative Results)

Qualitative analysis of the feedback revealed several key themes regarding the user experience and suggestions for improvement. Participants provided free-text responses regarding their VR relaxation task experience, which were analyzed for future technology improvement.

Immersion and Visual Artifacts (Improve Realism and Reduce Pixelation)

Some participants struggled with visual quality, stating that the graphics were "bland" and "pixelated." One participant mentioned, "The environment didn't feel real enough to help me relax."

Discomfort With the Headset (Select Lighter Weight Hardware)

Participants found the VR headset too heavy, making it difficult to use for prolonged relaxation. One user commented, "The headset was too bulky—it distracted me rather than helping me relax."

Voice Guidance Issues (Offer Customizable Audio Settings)

While some users appreciated the guided relaxation, others found the voiceover distracting or repetitive. One participant stated, "The voice instructions were too constant—I wanted more silence to focus on breathing."

Motion Sickness and Unpleasant Sensations (XR May Minimize Some Disorienting Effects)

A few participants experienced dizziness, with one stating, "The moving visuals made me feel nauseous, which completely defeated the point of relaxing." This suggests a need for less intense motion effects.

Mixed User Feedback on Effectiveness (Offer Alternatives, Eg, Audio-Only Modes)

Some participants felt the VR relaxation could be beneficial if improved, while others stated they would prefer alternative relaxation methods (eg, audio-only relaxation without VR).

Participants also provided free-text responses regarding their experience with the XR position feedback task, which were analyzed for future technology improvement.

Real-Time Visual Feedback Issues (Lower Latency Motion Tracking)

Some participants struggled with feedback clarity, reporting inconsistencies in motion tracking. One participant noted, "Sometimes my arm was perfectly aligned, but it wouldn't register the movement."

Difficulty in Adjusting Position (Online Recalibration)

A few participants found it difficult to match their movements with the system's feedback. One participant commented, "I kept missing the hoop even when I thought I was on target." Another commented, "I liked that it gave immediate feedback, but sometimes I didn't understand what I did wrong." This suggests that target alignment and hit detection need refinement.

Engagement and Gamification Elements (Expand Game-Like Elements)

Some participants enjoyed the interactive aspect of the task. One participant stated, "It was fun trying to score points, but I wish there were more levels or challenges."

Physical Strain Concerns (Individualized Task Intensity)

A small number of participants reported discomfort or strain during prolonged use. One participant mentioned, "I could feel my arm getting tired quickly—I think the tracking required more effort than I expected."

Mixed User Feedback on Usability (Lower Latency Motion Tracking and Online Recalibration)

Some participants felt that improving the accuracy and responsiveness of the tracking would make the task more engaging. One participant suggested, "If it was more precise in detecting movements, I'd find it much more enjoyable."

Participants provided additional free-text comments about their experience using the XR force feedback task, which were analyzed for future technology improvement.

Lack of Personalization (Individualized and Adaptive Resistance)

Several participants noted that the resistance levels were not well-adjusted to their needs. One participant stated, "The force applied felt either too weak or too strong—there was no in-between." This suggests a need for adaptive resistance control.

Discomfort and Fatigue (Improve Ergonomics)

The heaviness of the headset and the effort required to overcome force resistance were cited as major concerns. One participant reported, "After a few minutes, my arm felt very fatigued, which made the task frustrating rather than helpful." Another stated, "The device felt restrictive rather than supportive."

Low Engagement (Expand Game-Like Elements)

The lack of an interactive or gamified element was also highlighted. One participant commented, "There's no motivation to keep going—it's just moving against resistance with no real feedback."

Potential for Improvement (Future Potential)

Some participants saw promise in the concept but suggested improvements, such as, "It would help if the system guided me on whether I was applying the right force," "Maybe add vibration or a sound effect when I get the force correct," and "If the resistance could change based on how strong I am, that would be much better."

Summary

In summary, the qualitative feedback provided actionable information that complemented the SUS usability scores. It explains why certain tasks received lower scores (eg, VR relaxation's technical and content issues leading to poor ratings from half the group) and reinforces the need for customization in the force feedback task (given one user's difficulties). The participatory nature of the co-design and usability testing session ensured that end-user voices directly informed the next steps of platform refinement.

Discussion

Principal Results

This study is, to our knowledge, the first mixed methods evaluation of an XR-based biofeedback training platform co-designed for individuals with motor FND. Through a 2-phase coproduction approach, we obtained rich stakeholder input and preliminary evidence of usability. Our key finding is that the XR position feedback game was the most well-received component of the platform, with consistently high SUS usability scores and positive feedback from users. This task, dubbed "Hoop Hustle" in the prototype, required participants to perform wrist movements to control a VR interface with accompanying visual feedback. The strong performance of this task suggests that combining visual feedback in an intuitive pointing game can be highly engaging and easy to use for people with functional weakness. Participants likely benefited from the clear, immediate cause-and-effect in this game, which may have contributed to a sense of accomplishment and control.

In contrast, the VR relaxation module yielded a polarized reaction: some individuals felt deeply relaxed and enjoyed the experience (reflected in very high SUS scores), while others struggled with aspects of the VR environment (leading to poor usability ratings). These divergent outcomes highlight that a one-size-fits-all relaxation experience may not suit everyone; factors such as susceptibility to motion sickness, comfort with wearing a VR headset, and personal preference for meditation-style activities can greatly influence one's experience. The XR force feedback task showed intermediate and more variable usability. Most participants handled the force-feedback task moderately well (SUS~80‐95 for 4 participants), indicating that they understood the task and could perform it, but one participant (P3) had an extremely negative experience (SUS 27.5). P3's case is particularly informative: this participant has functional dystonia (a subtype of FND causing involuntary muscle contractions), which likely made it difficult to perform the steady force output required by the task. This resulted in frustration or fatigue, as reflected in both the low usability rating and the participant's comments describing the force task as "hard to manage" and "tiring." This finding underscores that individual clinical differences (such as the type of motor symptoms) can dramatically affect the usability of specific training tasks. Notably, the same participant (P3) rated the XR position task very highly (92.5), much higher than they rated the other 2 tasks. We interpret this to mean that while the force feedback task was not well-tolerated by P3, the position feedback game was accessible and enjoyable even for someone with dystonia. It is possible that the position task's design—emphasizing range of motion and coordination rather than sustained force—was better aligned with this participant's abilities. This suggests a need for personalized task selection or customization: users might benefit from having multiple training task options and skipping or modifying those that aggravate their symptoms.

Across all tasks, the qualitative feedback provided further insight into these quantitative results. For instance, participants who gave lower SUS scores often cited specific issues that explained their discomfort. Those who rated the VR relaxation poorly mentioned problems like visual graininess and a sense of disorientation when the virtual scene "breaks" (one user described a loss of immersion at a certain transition, eg, reaching a virtual staircase where the illusion was not convincing). On the other hand, participants who enjoyed the relaxation task commented on feeling calm and appreciating the break from active gameplay, which may reflect personal differences in how individuals prefer to engage (active interaction vs passive relaxation). Similarly, mixed feedback on the force task

corresponded with whether users felt the haptic feedback was appropriate; some found it novel and motivating, while others found it confusing or difficult to calibrate their strength. We can summarize the qualitative feedback themes as follows.

Hardware Comfort and Ergonomics

Multiple participants commented on the VR headset's weight and fit. One noted that the "headset is heavy" and that the straps were "a bit fiddly" to adjust properly. Another participant suggested the need for a more personalized or lightweight headset, saying they "would prefer [their] own personal headset" if using the system regularly. These comments indicate that physical comfort is a crucial factor, as discomfort could limit how long users with FND (who may have neck or upper body weakness) can wear the device. Ensuring a better fit and lighter hardware in future versions was a unanimous priority among participants.

Immersiveness and Visual or Auditory Feedback

Participants generally appreciated the concept of the immersive training tasks, but they pointed out specific issues that broke their sense of immersion. For instance, one participant observed that in the VR relaxation, "the picture quality is bland" (low resolution), which detracted from the experience. Visual artifacts or graphics glitches were noticed by another, who commented that such issues "break immersion." On the auditory side, a few participants felt the guided meditation voice-over in the VR relaxation was "too artificial" and constant, making it "distracting" rather than soothing. One user recommended incorporating periods of silence or softer, nonverbal audio, noting that "Constant speech is too much–needs time to breathe." In the XR game, participants enjoyed the sound effects, but one suggested adding more varied sound cues for feedback (eg, different sounds when a hoop is scored versus missed). Enhancing the realism and quality of sensory feedback (both visual and auditory) would likely improve user engagement.

Task Difficulty and Personalization

There was a strong consensus on the importance of adjusting the tasks to individual capabilities. In the hoop hustle game, participants had different skill levels; 1 patient with a more severe weakness struggled initially, so the facilitator enlarged the hoop and reduced the required movement range. This kind of on-the-fly personalization was appreciated. Participants explicitly mentioned features they would like to see: "adjustable height [of hoops]" and "hoop size" options, as well as the ability to slow down or speed up the game pace. In the force feedback task, the participant with dystonia noted that the resistance made the task quite challenging for them, but felt it might be helpful if it could be tuned to their strength level. Across the feedback, "personalization" emerged as a key theme–one size does not fit all in this diverse group. Future versions of the platform should include user-specific calibration, difficulty settings, and possibly adaptive algorithms that modify task parameters in real-time based on performance.

Perceived Benefits and Engagement

Despite the critiques, most participants expressed enthusiasm for the platform's concept. Several referred to the approach as a "brilliant idea" and were eager to see it refined. They reported finding the interactive game enjoyable–one health care professional noted that the competitive element of trying to get the ball through the hoop "made it fun, so you forget you're exercising." Participants also believed the platform could increase patient motivation to perform rehabilitation exercises, as it "doesn't feel like therapy" in the traditional sense. The relaxation task was seen as potentially useful for calming down patients before or after physical exercises, although it clearly needs improvement to be effective for everyone.

Practical Considerations (Accessibility)

Echoing the Delphi survey results, workshop participants raised practical questions. They debated whether the system would be used in clinics or at home. For home use, participants stressed the need for proper guidance and support: "If this was sent to patients, there would need to be a help guide or 24/7 tech support," one participant said, concerned about less tech-savvy users. The idea of a shared device versus personal ownership was discussed; some felt a single headset could be used by multiple patients in a clinic if properly sanitized, while others thought long-term users would benefit from having their own device configured to their needs. Concerns about cost were mentioned again; one participant estimated "it's [£]1000… (US $1330) I could not afford [this]" and hoped it would be provided through the NHS or insurance. These discussions highlight that for the platform to be implementable, issues of cost, training, and technical support must be addressed alongside its technical development.

These thematic insights demonstrate the value of a mixed methods approach: the quantitative data identified where usability was strong or weak, and the qualitative data helped explain why those outcomes occurred. Crucially, the workshop confirmed that co-design is not only feasible but beneficial in developing neurotechnology for FND. Participants' real-time feedback led us to identify specific improvements (eg, modifying the VR content and adding adjustable settings in the game) that we might not have fully appreciated without their involvement. The inclusion of both patients and clinicians ensured that the usability assessment considered practical use in a clinical context.

Comparison with Previous Work

Our findings align with existing literature emphasizing user-centered design for health technologies. Previous studies have noted that even small samples (5 - 10 users) can uncover the majority of usability issues in a system [29]. In our case, 6 users were sufficient to highlight distinct strengths and weaknesses of the platform. The variability in VR relaxation feedback is reminiscent of observations in broader VR applications: while VR can provide immersive therapeutic experiences, factors like motion sickness and comfort remain challenges to address. The need for personalization in rehabilitation technology is well-documented; for instance, usability studies of other rehab games have found that adaptive difficulty can significantly improve user engagement and outcomes. Our results specifically extend this understanding to FND, suggesting that personalization may not only improve engagement but might be necessary to accommodate neurological symptoms like dystonia or fatigue. From a

neurological perspective, the concept of using haptic feedback and VR to retrain the perception-action cycle in FND draws on theories of sensory attenuation and agency in functional movement disorders. By providing congruent visual and haptic inputs corresponding to the user's intended movements, the platform aims to reinforce the association between effort and sensory feedback, potentially strengthening the efference copy mechanism that is hypothesized to be underactive in patients with FND [14,30,31]. While our study did not directly measure clinical outcomes or neurophysiological changes, the positive usability of the position and force feedback tasks is a critical first step toward implementing such therapeutic concepts in practice. A recent review by [12] also emphasized VR's promise for addressing mechanisms of agency and attention in FND; our practical findings complement this by showing that patients are willing to engage with VR or haptic systems, provided they are comfortable and accessible.

Limitations

This study has several limitations. First, the sample size was small (5 - 6 participants for usability testing), and all participants were from a single clinical center and also coauthors, which could introduce some bias or limit critical feedback. The findings should be interpreted as preliminary and exploratory; a larger, independent sample will be needed to validate and generalize the usability results. Second, participants' familiarity with XR technology varied, and those with previous VR or gaming experience might have found the system easier to use, potentially influencing their SUS scores. We did not formally quantify each participant's XR technology background, which is a confounding factor that future studies should measure. Third, we focused on 3 specific XR tasks (VR relaxation, XR position feedback, and XR force feedback). Other functionalities (eg, bilateral training or cognitive tasks in XR) were not included and could present additional usability challenges or benefits not captured here. Fourth, the reliance on subjective SUS scores introduces potential bias, as individual expectations or novelty effects can influence ratings. We mitigated this by collecting detailed qualitative feedback, but objective performance metrics were not analyzed in this pilot. Fifth, as an initial co-design and usability study, we did not assess clinical efficacy, for instance, whether using the platform yields improvements in motor function or FND symptoms. Such outcomes will need evaluation in subsequent trials. Finally, our personalization of the tasks was done manually by the facilitators rather than through built-in adaptive algorithms. This limits the consistency of the user experience; an automated personalization mechanism would be ideal to ensure each user gets an optimally challenging experience. Despite these limitations, the study provides valuable insights into the user experience of an XR neurotechnology platform tailored for FND. To our knowledge, this is one of the first studies to report detailed usability data for an XR haptics platform in FND rehabilitation. The co-design approach proved effective in identifying user priorities and potential pitfalls early in the development process.

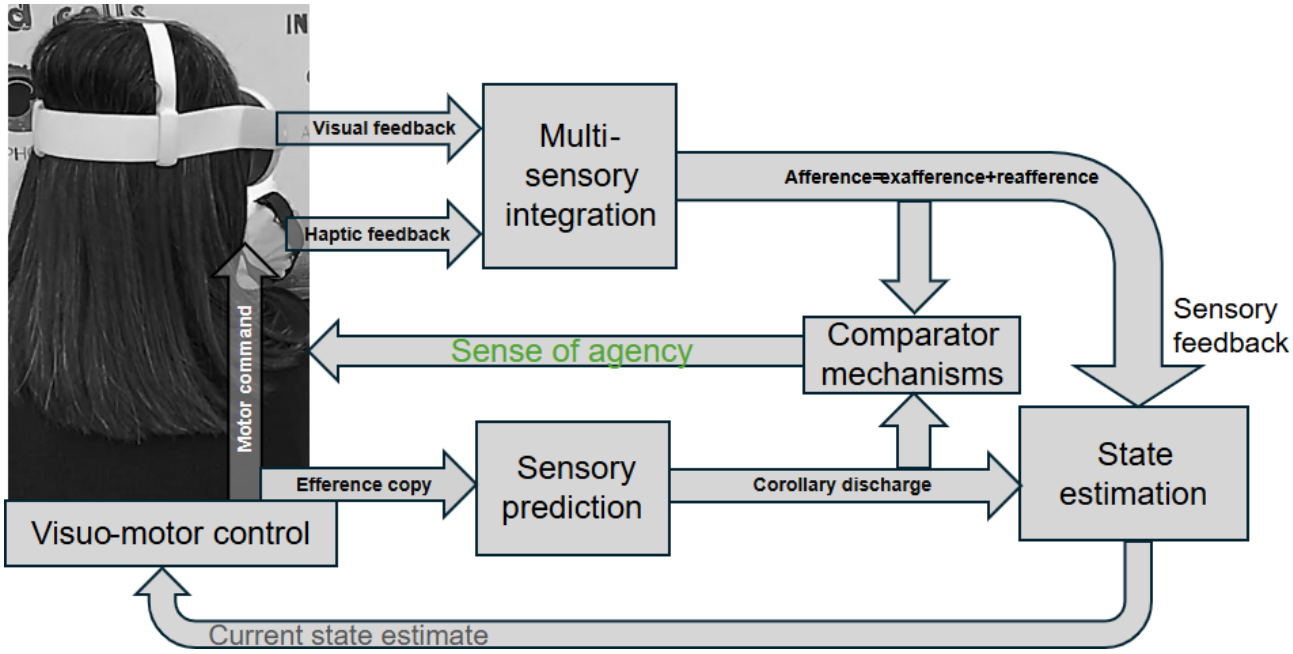
Future Directions

The next steps following this study will address the identified issues and test the platform on a broader scale at home [32]. FMD involves involuntary-feeling but voluntary-appearing movements, linked to a disrupted sense of agency due to impaired sensory attenuation, that is, the brain's predictive suppression of self-generated sensory feedback [33]. This impairment, involving brain regions like the primary motor cortex, cerebellum, and right temporo-parietal junction, leads to difficulties in distinguishing self-initiated actions from external stimuli [34]. Conversely, sensory amplification, mediated by the posterior parietal cortex, heightens sensory perception through attention. XR presents therapeutic potential by balancing sensory attenuation and amplification [6]. VR allows controlled manipulation of predictive coding, helping recalibrate agency and sensory processing in FMD. The comparator model (refer to Figure 3) suggests that agency arises when predicted sensory outcomes align with actual feedback, which can be reinforced through haptic feedback in XR. Here, linking active inference in motor control lies in its ability to explain and address motor dysfunctions [35]. By recognizing that the brain updates perceptions and modifies actions to minimize prediction errors, this framework offers insights into abnormal motor control, where disrupted sensory prediction leads to impaired agency and movement errors. In rehabilitation, this perspective supports the development of XR biofeedback interventions, where haptic and visual feedback can help recalibrate faulty sensorimotor predictions. By reinforcing accurate sensory-motor associations, these technologies may restore agency and improve motor function, offering a novel, personalized approach to therapy. Indeed, XR technologies have been shown to enhance sensorimotor processing, but usability for patients with FMD must be assessed. Early user involvement, particularly in conditions like functional dystonia, is critical to refining XR rehabilitation design. Our study engaged stakeholders from academia, industry, and health care (NHS, England) to classify technological needs into incremental or revolutionary advancements. Notably, no commercial or research-based XR biofeedback systems currently exist specifically for FMD rehabilitation.

In response to user feedback, we are working with the developers to improve the VR relaxation module (enhancing graphics, refining the audio guidance, and possibly adding options for different scenes or background music) [32]. We are also implementing in-software settings that allow end-users or therapists to easily adjust game difficulty, visual or auditory feedback levels, and force feedback intensity. In addition, we plan to incorporate a brief calibration or tutorial at the start of a session, where the system can gauge a user's comfortable range of motion and strength, and automatically set initial task parameters accordingly. These changes aim to embed personalization directly into the platform. A follow-up study is being designed to involve a larger cohort of patients with FND in a multisession at-home trial with the refined platform [32]. That study will evaluate not just usability, but also short-term effects on motor function and symptoms, using clinical scales and objective performance metrics within the game. We will also examine learning effects–whether repeated use leads to improved user proficiency or changes in feedback preferences–to understand how usability evolves over time. An important future direction is to explore remote or home usability of this platform [32]. Given the interest in home-based rehabilitation (and lessons

learned during the COVID-19 pandemic), we aim to test whether patients can effectively use the XR system at home with minimal supervision. This will involve developing comprehensive user guides, integrating remote monitoring capabilities (so therapists can track usage and progress), and ensuring robust technical support is available. Addressing these factors will be essential for translating this coproduced XR platform into a scalable, real-world therapeutic option for individuals with FND.

Figure 3. Based on the comparator model, when a motor command is issued, an accompanying efference copy is generated, which allows the brain to predict the expected sensory outcome of the action. This predicted outcome is then compared to the actual sensory feedback upon action completion. A strong the feeling of agency is experienced if there is a close match between predicted (corollary discharge) and actual sensory information (afference) from the environment. This comparator model can also explain feeling of agency in virtual extended reality (XR) environments where a virtual representation mimics the user's physical movement, providing exafference that, when combined with reafference, provides users the sense of agency (feeling of agency).



Conclusions

Through a collaborative coproduction approach, we developed and pilot-tested a novel XR (VR+ haptic) biofeedback training platform for patients with functional upper limb weakness due to FND. Our usability findings are encouraging: an interactive XR position feedback game was rated highly usable by all participants, and a VR relaxation experience received very positive feedback from some users. At the same time, the variability in responses, particularly the challenges faced by one participant during the force feedback task, highlights the necessity of a flexible, user-tailored design in such neurotechnologies. One-size-fits-all solutions are unlikely to succeed in the FND population given the diversity of symptoms and user preferences. By systematically incorporating user feedback, we identified concrete areas for improvement (such as hardware comfort and software adaptability) that will guide the next iteration of the platform. This study demonstrates that patients with FND and clinicians are not only capable of providing meaningful input into technology design but are eager to do so when the goal is to enhance therapy. With further refinement and larger-scale testing, the XR platform has the potential to become a valuable tool in FND rehabilitation, offering engaging, at-home training that reinforces patients' agency and motor function in a way that is enjoyable and customized to their needs.

Acknowledgments

The authors extend their sincere thanks to all participants for their time and invaluable feedback. We gratefully acknowledge funding support from the N-CODE EPSRC Network+ for the hot-topic workshop on the cocreation of digital health technologies for the management of Functional Neurological Disorders held on October 11, 2023 at the University of Central Lancashire, UK. The EPSRC Rehabilitation Technologies Network+ funded this project in 2024 on the coproduction of the extended reality (XR) biofeedback platform. Preliminary findings were presented by Matthew Alexander Newsham at the National Rehabilitation Technologies Conference 2024 (September 17–18, 2024) and by Katerina Hatjipanagioti at the Royal Society's Digital Healthcare for the Management of Functional Neurological Disorders symposium (November 25–26, 2024). Feedback from all these events was instrumental in shaping the development of this work. We also acknowledge the international recruitment support provided by members of the re+active FND Membership program, which facilitated participation in the online Delphi survey.

Authors' Contributions

Dutta A was responsible for conceptualization, lead funding acquisition, lead methodology development, visualization, original draft writing, and contributed equally to project administration and supervision. LR, KH, and MN contributed equally to data curation and investigation. LL, JT, and Das A also contributed equally to the investigation. AB, IF, and Das A provided supporting roles in funding acquisition and methodology. JT led in providing resources, with AB and IF offering supporting contributions. Das A performed validation and contributed equally with Dutta A to project administration and supervision. For writing—review and editing, Dutta A took the lead, with supporting contributions from Das A, AB, IF, and JT.

Conflicts of Interest

AB is the CEO of Nudge Reality Ltd., and IF is the CEO of Human Robotix Ltd. Both provided technology and expertise for this project. Their involvement was limited to technical development, and they were not involved in the analysis of usability data. The other authors declare no competing interests.

Multimedia Appendix 1
NRC Rehabilitation Technologies Conference 2024 poster.
[PDF File, 2895 KB - xr_v2i1e68580_app1.pdf]

Multimedia Appendix 2
NRC Rehabilitation Technologies Conference 2024 slides.
[PDF File, 928 KB - xr_v2i1e68580_app2.pdf]

Multimedia Appendix 3
Human Robotix's HRX-1 system.
[PDF File, 655 KB - xr_v2i1e68580_app3.pdf]

Multimedia Appendix 4
Nudge Reality's XR games.
[PDF File, 131 KB - xr_v2i1e68580_app4.pdf]

Multimedia Appendix 5
XR System Usability testing script.
[PDF File, 42 KB - xr_v2i1e68580_app5.pdf]

Multimedia Appendix 6
XR System Usability testing results – Force Feedback.
[ZIP File, 12330 KB - xr_v2i1e68580_app6.zip]

Multimedia Appendix 7
XR System Usability testing results – PositionFeedback.
[ZIP File, 12241 KB - xr_v2i1e68580_app7.zip]

Multimedia Appendix 8
XR System Usability testing results – Relaxation.
[ZIP File, 10073 KB - xr_v2i1e68580_app8.zip]

References

1. Dutta A. Bayesian predictive coding hypothesis: Brain as observer's key role in insight. Med Hypotheses 2025 Feb;195:111546. [doi: 10.1016/j.mehy.2024.111546]
2. O'Mahony B, Nielsen G, Baxendale S, Edwards MJ, Yogarajah M. Economic Cost of Functional Neurologic Disorders: A Systematic Review. Neurology (ECronicon) 2023 Jul 11;101(2):e202-e214. [doi: 10.1212/WNL.0000000000207388] [Medline: 37339887]
3. Hallett M, Aybek S, Dworetzky BA, McWhirter L, Staab JP, Stone J. Functional neurological disorder: new subtypes and shared mechanisms. Lancet Neurol 2022 Jun;21(6):537-550. [doi: 10.1016/S1474-4422(21)00422-1] [Medline: 35430029]
4. FND Hope International. Freedom of information project aimed to explore issues reported by NHS clinicians and the FND community regarding accesses to treatment for people with FND. URL: https://www.fndhope.org.uk/

freedom-of-information-project-aimed-to-explore-issues-reported-by-nhs-clinicians-and-the-fnd-community-regarding-accesses-to-treatment-for-people-with-fnd/ [accessed 2024-11-09]
5. Taing M, Sztainert T, Harley M, Allen K, Moore F. Applying a Stepped-Care Framework for Functional Neurological Disorder Management. Can J Neurol Sci 2025 Jan;52(1):39-41. [doi: 10.1017/cjn.2024.32]
6. Dutta A, Das A. A platform technology for VR biofeedback training under operant conditioning for functional lower limb weakness. Engineering. Preprint posted online on 2024. [doi: 10.20944/preprints202407.2520.v1]
7. Kumar D, Sinha N, Dutta A, Lahiri U. Virtual reality-based balance training system augmented with operant conditioning paradigm. Biomed Eng Online 2019 Aug 28;18(1):90. [doi: 10.1186/s12938-019-0709-3] [Medline: 31455355]
8. Ducroizet A, Zimianti I, Golder D, et al. Functional neurological disorder: Clinical manifestations and comorbidities; an online survey. J Clin Neurosci 2023 Apr;110:116-125. [doi: 10.1016/j.jocn.2023.02.014] [Medline: 36871491]
9. Stone J, Carson A, Hallett M. Explanation as treatment for functional neurologic disorders. Handb Clin Neurol 2016;139:543-553. [doi: 10.1016/B978-0-12-801772-2.00044-8] [Medline: 27719870]
10. Hirjak D, Fritze S, Northoff G, Kubera KM, Wolf RC. Die sensomotorische Domäne im Research-Domain-Criteria-System: Fortschritte und Perspektiven. Nervenarzt 2021 Sep;92(9):915-924. [doi: 10.1007/s00115-021-01144-7]
11. Spagnolo PA, Garvey M, Hallett M. A dimensional approach to functional movement disorders: Heresy or opportunity. Neurosci Biobehav Rev 2021 Aug;127:25-36. [doi: 10.1016/j.neubiorev.2021.04.005] [Medline: 33848511]
12. Brouwer D, Morrin H, Nicholson TR, et al. Virtual reality in functional neurological disorder: a theoretical framework and research agenda for use in the real world. BMJ Neurol Open 2024;6(2):e000622. [doi: 10.1136/bmjno-2023-000622] [Medline: 38979395]
13. Kilteni K, Andersson BJ, Houborg C, Ehrsson HH. Motor imagery involves predicting the sensory consequences of the imagined movement. Nat Commun 2018 Apr 24;9(1):1617. [doi: 10.1038/s41467-018-03989-0] [Medline: 29691389]
14. Dutta A. “Hyperbinding” in functional movement disorders: role of supplementary motor area efferent signalling. Brain Commun 2025;7(1):fcae464. [doi: 10.1093/braincomms/fcae464] [Medline: 39995655]
15. Wolpert DM, Miall RC. Forward Models for Physiological Motor Control. Neural Netw 1996 Nov;9(8):1265-1279. [doi: 10.1016/s0893-6080(96)00035-4] [Medline: 12662535]
16. Ebner TJ. Cerebellum and internal models. In: Handbook of the Cerebellum and Cerebellar Disorders: Springer; 2013:1279-1295. [doi: 10.1007/978-94-007-1333-8_56]
17. Centre for Applied Education Research. A synthesis of the many errors and learning processes of visuomotor adaptation. URL: https://caer.org.uk/publications/26ec6253606a6d05e543c56eb179de70b6fe6c34/ [accessed 2024-11-09]
18. Müller T, Klein-Flügge MC, Manohar SG, Husain M, Apps MAJ. Neural and computational mechanisms of momentary fatigue and persistence in effort-based choice. Nat Commun 2021 Jul 28;12(1):4593. [doi: 10.1038/s41467-021-24927-7] [Medline: 34321478]
19. Moore JW, Obhi SS. Intentional binding and the sense of agency: a review. Conscious Cogn 2012 Mar;21(1):546-561. [doi: 10.1016/j.concog.2011.12.002] [Medline: 22240158]
20. Stone J, Edwards M. Trick or treat? Showing patients with functional (psychogenic) motor symptoms their physical signs. Neurology (ECronicon) 2012 Jul 17;79(3):282-284. [doi: 10.1212/WNL.0b013e31825fdf63] [Medline: 22764261]
21. Das A, Newsham M, Hatjipanagioti K, Buchanan A, Farkhatdinov I, Dutta A. VR haptics biofeedback training for functional limb weakness: insights from the first round of a delphi survey. Presented at: National Rehabilitation Centre’s Conference on Rehabilitation Technologies; Sep 17-18, 2024; University of Nottingham URL: https://www.researchgate.net/publication/382686730_VR_Haptics_Biofeedback_Training_for_Functional_Limb_Weakness_Insights_from_the_First_Round_of_a_Delphi_Survey
22. Cecamore M, Gaponov I, Miller SC, Farkhatdinov I. Design and validation of a haptic ankle platform for physical human-machine interaction. Presented at: 2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob); Sep 1-4, 2024; Heidelberg, Germany p. 1301-1306. [doi: 10.1109/BioRob60516.2024.10719789]
23. Berret B, Verdel D, Burdet E, Jean F. Co-contraction embodies uncertainty: An optimal feedforward strategy for robust motor control. PLOS Comput Biol 2024 Nov;20(11):e1012598. [doi: 10.1371/journal.pcbi.1012598] [Medline: 39565821]
24. Cazenave L, Einenkel M, Yurkewich A, Endo S, Hirche S, Burdet E. Hybrid Robotic and Electrical Stimulation Assistance Can Enhance Performance and Reduce Mental Demand. IEEE Trans Neural Syst Rehabil Eng 2023;31:4063-4072. [doi: 10.1109/TNSRE.2023.3323370] [Medline: 37815973]
25. Ivanova E, Eden J, Carboni G, Krüger J, Burdet E. Interaction with a reactive partner improves learning in contrast to passive guidance. Sci Rep 2022 Sep 22;12(1):15821. [doi: 10.1038/s41598-022-18617-7] [Medline: 36138031]
26. John B. SUS -- a Quick and Dirty Usability Scale: Usability Evaluation In Industry; 1996.
27. Bangor A, Kortum PT, Miller JT. Determining what individual SUS scores mean: adding an adjective rating scale. JUX - The Journal of User Experience 2009(3):114-123. [doi: 10.5555/2835587.2835589]
28. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;14(1):e8. [doi: 10.2196/jmir.2042]
29. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]

30. Pringsheim T, Edwards M. Functional movement disorders: Five new things. Neurol Clin Pract 2017 Apr;7(2):141-147. [doi: 10.1212/CPJ.0000000000000350] [Medline: 29185540]
31. Pareés I, Brown H, Nuruki A, et al. Loss of sensory attenuation in patients with functional (psychogenic) movement disorders. Brain (Bacau) 2014 Nov;137(Pt 11):2916-2921. [doi: 10.1093/brain/awu237] [Medline: 25161293]
32. Platform technology for VR biofeedback training under operant conditioning for functional limb weakness: A proposal for co-production of at-home solution (REACT2HOME). JMIR Preprints. URL: https://preprints.jmir.org/preprint/70620 [accessed 2025-03-02]
33. Haggard P, Whitford B. Supplementary motor area provides an efferent signal for sensory suppression. Brain Res Cogn Brain Res 2004 Mar;19(1):52-58. [doi: 10.1016/j.cogbrainres.2003.10.018] [Medline: 14972358]
34. Zito GA, Wiest R, Aybek S. Neural correlates of sense of agency in motor control: A neuroimaging meta-analysis. PLOS ONE 2020;15(6):e0234321. [doi: 10.1371/journal.pone.0234321] [Medline: 32502189]
35. Brown H, Adams RA, Parees I, Edwards M, Friston K. Active inference, sensory attenuation and illusions. Cogn Process 2013 Nov;14(4):411-427. [doi: 10.1007/s10339-013-0571-3] [Medline: 23744445]

Abbreviations

FMD: functional motor disorder
FND: functional neurological disorder
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
HRX-1: human robotix system
MR: mixed reality
NHS: National Health Service
NRC: National Rehabilitation Centre
PPIE: Patient and Public Involvement and Engagement
SUS: system usability scale
VR: virtual reality
XR: extended reality

Edited by K Martinez, L Riedemann; submitted 09.11.24; peer-reviewed by J Brooke, X Cheng; revised version received 02.03.25; accepted 15.03.25; published 03.06.25.

Please cite as:
Dutta A, Hatjipanagioti K, Newsham MA, Leyland L, Rickson L, Buchanan A, Farkhatdinov I, Twamley J, Das A
Extended Reality Biofeedback for Functional Upper Limb Weakness: Mixed Methods Usability Evaluation
JMIR XR Spatial Comput 2025;2:e68580
URL: https://xr.jmir.org/2025/1/e68580
doi:10.2196/68580

Augmented Reality Framework to Measure and Analyze Eye–Hand Coordination in Stroke Patients with Unilateral Neglect: Proof-of-Concept Study

Jonathan Becker[1], MSc; Sophokles Ktistakis[1], MSc; Mirko Meboldt[1], PhD; Thomas Nyffeler[2,3], MD; Dario Cazzoli[2,3], PhD; Quentin Lohmeyer[1], PhD

[1]ETH Zurich, Leonhardstrasse 21, Zurich, Switzerland

[2]Neurocenter, Lucerne Cantonal Hospital, University Teaching and Research Hospital, University of Lucerne, Lucerne, Switzerland

[3]Faculty of Behavioural Sciences and Psychology, University of Lucerne, Lucerne, Switzerland

Corresponding Author:
Jonathan Becker, MSc
ETH Zurich, Leonhardstrasse 21, Zurich, Switzerland

Abstract

Background: Stroke is a leading cause of disability, often accompanied by unilateral spatial neglect (USN), which severely impairs recovery. Traditional assessments like paper-pencil tests provide limited insights into behaviors and eye–hand coordination during real-world tasks. Advances in hand pose estimation and eye tracking in combination with augmented reality (AR) offer potential for data-driven assessments of naturalistic interactions.

Objective: This proof-of-concept study presents and evaluates a multimodal behavioral tracking system that captures gaze, body, and hand movements during interactions within an AR environment. Our primary goals are to (1) validate that this system can achieve robust and accurate interaction data capture in clinical settings, (2) show that the system can reliably detect known USN behavioral patterns, and (3) explore how comprehensive data can provide new understanding of eye–hand coordination deficits in USN.

Methods: We developed an AR-based assessment system using Microsoft HoloLens 2 and an external body-tracking camera to capture real-time gaze, hand, and body movements in an interactive environment. Multimodal data streams were temporally synchronized, fused, and filtered to enhance spatial accuracy and availability. Tracking performance was benchmarked against a traditional optical motion-capture system to validate reliability. In a study, 7 patients with right-brain lesions with mild to moderate USN and 8 healthy controls participated. Each performed a designed reaching task, stamping virtual sheets of paper that appeared randomly on a table. We analyzed participants' search behavior patterns to assess attentional biases and examined gaze anchoring timing during targeted reaching motions to explore potential eye–hand coordination deficits.

Results: The fusion of hand-tracking data from the HoloLens 2 and external system reduced tracking loss from 25.7% to 2.4%, with an absolute trajectory error of 3.27 cm. The system demonstrated high usability and was well accepted by patients. Data from the control group confirmed the absence of intrinsic lateral biases in the system and task design. The USN group displayed typical search behavior through ipsilesional biases in gaze direction during visual exploration (median deviation 7.46 [1.61-9.48] deg, $P<.05$) and longer times to find contralesional targets (median difference 1.08 [0.20-1.80] s, $P=.02$). Additionally, the eye–hand coordination analysis revealed lateral differences in gaze anchoring during targeted reaching motions in the USN group, with earlier fixation on contralesional targets (median difference 112 [71-146] ms, $P=.02$).

Conclusions: The proposed AR framework provides a novel, comprehensive data-driven method for capturing interaction behavior in a controlled, yet naturalistic environment. Our results demonstrated the system's effectiveness in measuring hallmark USN symptoms, such as gaze and head orientation biases, and highlighted its potential to complement traditional assessments by offering deep insights into torso rotation and eye–hand coordination with a high resolution and accuracy. This data-driven approach shows promise for enhancing current USN assessment practices and gaining new insights into patients' behaviors.

(JMIR XR Spatial Comput 2025;2:e70985) doi:10.2196/70985

KEYWORDS

visual neglect; stroke rehabilitation; augmented reality; eye-hand coordination; gaze anchoring; hand tracking; eye tracking; digital twin; motion capture; neurorehabilitation technology; unilateral spatial neglect; AR

Introduction

Stroke represents a profound public health challenge, affecting one in four adults during their lifetimes [1]. Of the numerous poststroke impairments, unilateral spatial neglect (USN) is a particularly significant barrier to patient recovery: Evident in up to 43% of patients after a right-hemispherical brain lesion (RBL) [2], this condition is characterized by a reduced ability to attend to contralesional stimuli, impacting essential daily activities [3,4]. Consequently, the rehabilitation journey is fraught with challenges, including heightened fall risks, poorer rehabilitation outcomes, extended hospitalization time, and reduced likelihood of being discharged home [5-7].

In 75% of USN cases, patients present with egocentric visual USN [8], marked by a bias of visual exploration through head and eye movements toward the ipsilesional side [9]. However, USN is a heterogeneous condition, encompassing a spectrum of attentional deficits across visual, tactile, auditory, and proprioceptive domains [10-12]. Consequently, multiple tests are typically employed to measure the severity of USN. Paper-pencil tests such as the Bells/Stars Cancelation [13] and Line Bisection [14] are commonly used due to their good sensitivity among paper-pencil assessments [15]. However, discrepancies between test results and patients' real-world behavior are frequently reported [16].

Behavioral assessments like the Catherine Bergego Scale address this gap by evaluating neglect during several activities of daily life, but they require labor-intensive direct observation and offer limited quantitative insight into task execution [17]. Technological approaches, such as gaze analysis during free visual exploration, provide quantitative insights but are restricted to specific test settings [18]. Evidence suggests that attentional biases in USN extend beyond eye movements, being influenced by motor acts and proprioceptive feedback. For instance, the orientation of the trunk mid-sagittal plane was shown to modulate the perceived forward direction and shift neglected space [19-22]. Moreover, proprioceptive cues, such as pointing, guide the direction of visual attention [23]. Accordingly, patients with USN exhibited biases of visual attention toward the ipsilesional sides of these cues [24].

These findings highlight the need for a comprehensive behavioral tracking system that can capture eye, body, and hand movements during task execution in environments that closely mimic real-world conditions.

Virtual reality (VR) or augmented reality (AR) systems that are able to track gaze and hand movements partially address these requirements. VR has been explored to simulate real-world interactions for studying USN [25]. However, immersive VR introduces new challenges such as user disconnection from the real environment, proprioceptive disconnection (eg, the inability to see one's own hands), and motion sickness [26]. By contrast, AR overlays simulated elements onto the real world, creating a mixed reality experience, addressing aforementioned discomforts [26] and integrating both physical and virtual objects into experiments. Consequently, AR has recently seen a growing application in neuroscience and rehabilitation [27,28].

However, despite these advantages, standalone AR systems still fall short of the comprehensive behavioral tracking requirements identified earlier. The inside-out tracking approach of typical AR systems like the Microsoft HoloLens 2 (where sensors on the headset track the environment) provides limited full-body tracking beyond hands and head position. Additionally, hand tracking can fail when users' hands are not positioned directly in front of them or during complex object interactions.

To address this gap, researchers have begun complementing inside-out tracking with external motion capture (Mocap) systems [29,30]. In clinical settings, external tracking systems combined with AR have primarily been used to position virtual objects relative to physical ones, particularly for surgical guidance applications [31,32]. However, there is limited evidence of this combination being utilized to capture comprehensive behavioral data in clinical rehabilitation contexts [33].

Accordingly, we propose a novel framework encompassing a head-mounted AR system (HoloLens 2, Microsoft USA [34]) and a single RGB-D camera (ZED 2i, Stereolabs USA), providing the following:

- A controlled environment allowing for a wide range of user interactions with both digital and real-world elements.
- Accurate tracking of gaze orientation, hand movements, and body posture during task execution, with improved hand-tracking accuracy and availability enabled by sensor fusion.
- Detailed digital reconstruction of interactions to enable comprehensive behavioral analysis.
- Automated data integration and metric extraction to enable quantitative analysis, with a focus on eye–hand coordination.

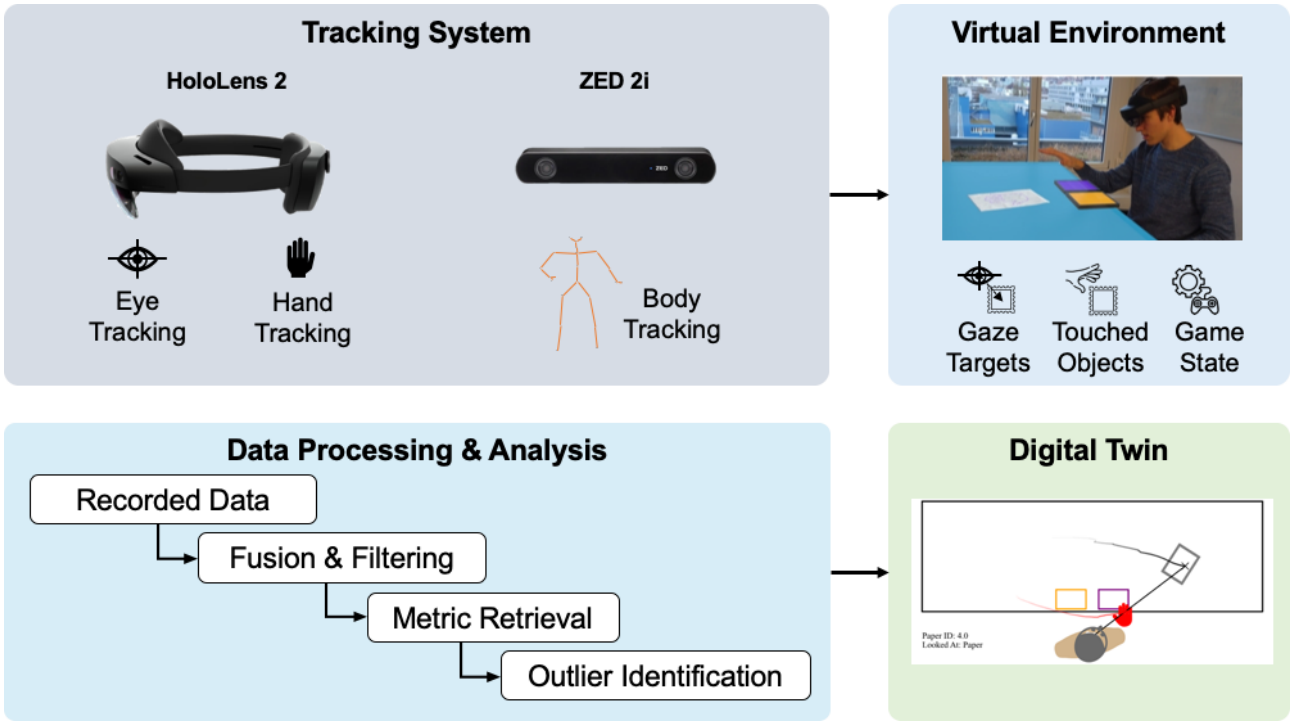
The goal of this proof-of-concept study is to demonstrate the feasibility and utility of our tracking framework through a systematic validation approach. Our validation encompasses both a technical accuracy assessment against a marker-based VICON system and clinical validation through the detection of known USN symptoms. Beyond confirming that the framework captures established USN behavioral patterns, we explore lateral differences in eye–hand coordination during goal-directed actions as a novel behavioral marker. With this, our framework shows promise in enhancing current USN assessment practices and gaining new insights into patients' behaviors.

Methods

System Overview

The system (Figure 1) comprised three main parts. (1) A tracking system, which measures and records gaze direction, hand, and body movements. It enables interactions with (2) a simulated virtual environment, which is overlaid onto the real world using a head-mounted AR display. (3) An offline data-processing and data-reconstruction system creates a digital twin of the captured interactions, allowing for in-depth analysis.

Figure 1. Overview of the system.



Tracking System

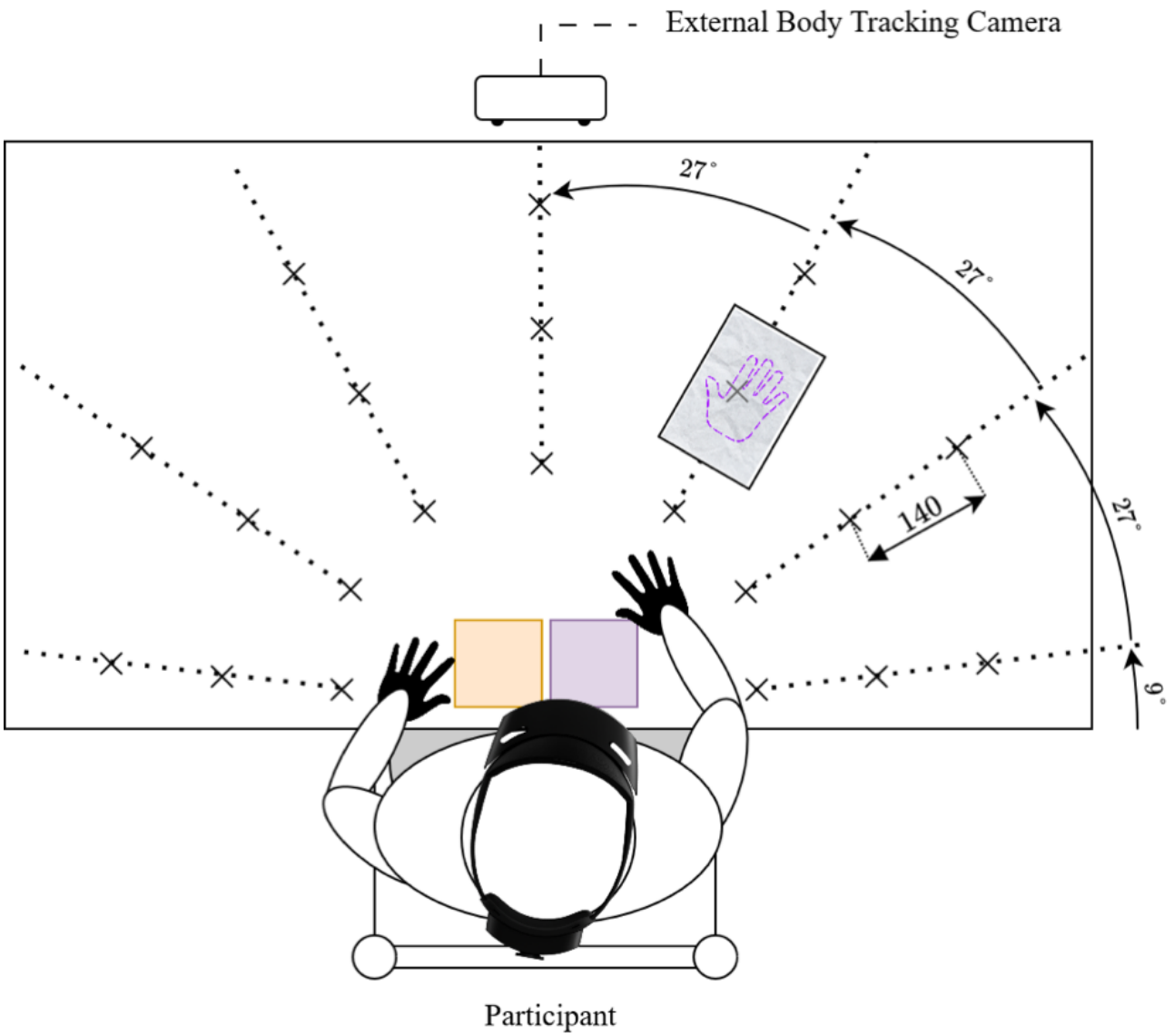
We obtained interaction data from multiple sources. The HoloLens 2 provided gaze and hand-tracking data through the Mixed Reality Toolkit API [35]. These data were recorded alongside the state of the virtual environment, at a frequency of 40 Hz, capturing the users' view at each moment. Additionally, an external RGB-D camera (ZED 2i, Stereolabs, USA) was set up in front of the users, facing them.

The coordinate frame was aligned to that of the HoloLens during setup by taking a picture of a checkerboard with both cameras and extracting the relative transformation between the two. Using the manufacturer-provided body-tracking algorithms [36], body key points were extracted, including hand and shoulder joint positions. This pose information was recorded at a frequency of 30 Hz. The hand-tracking data were additionally streamed in real time to the HoloLens to complement the on-device hand tracking that was at times subject to tracking loss.

Virtual Environment Design

The virtual environment was designed to assess the hand- and eye-tracking capabilities of our system, specifically its ability to capture metrics relevant to USN. To encourage naturalistic behavior, it was structured as a game around a familiar, easy-to-learn task that required repeated, precise reaching motions: stamping virtual sheets of paper. For this task, users sat in front of a physical table overlaid with a holographic table. Two virtual ink pads, one orange and one violet, were shown to participants, positioned as shown in Figure 2. During the task, virtual sheets of paper appeared sequentially in a random order across 21 locations distributed radially (Figure 2). Each sheet displayed a hand outline, color-coded in orange or violet. The users selected the color by putting their hand into the corresponding virtual ink pad in front of them. They then performed a reaching motion to stamp the paper, after which it disappeared, and a new one appeared. Participants were told that their main goal was to perform this task as fast as possible, but to ensure goal-oriented reaching and precise movements toward the targets, they were also told to stamp the papers as precisely as possible within the respective hand outline.

Figure 2. Top-down view of the game setup. Virtual sheets of paper appear in 21 possible locations indicated by X, radially distributed and spaced at 140 mm by default. Two virtual stamp pads are always shown in front of the user. The ZED camera for external body tracking is set up opposite of the user with a view of the entire table and the user's pose. Image of HoloLens 2 from [34].



To promote naturalistic behavior while ensuring data comparability, implicit constraints were implemented. A brief, randomized delay of 1 to 3 s after each paper was stamped encouraged users to reset their posture between trials. Moreover, the virtual sheets never appeared adjacent to each other, that is, as direct neighbors in any direction, to minimize carryover effects from the previous location. By requiring ink selection prior to stamping, each movement began from a central position and interactions followed a consistent sequence: locating the paper, selecting the ink, and finally stamping. These constraints ensured the independence and comparability of each episode without explicitly directing user actions.

An initial tutorial was introduced to help participants understand the game and the interaction with the system. This allowed users to familiarize themselves with the task and the limited field of view of the HoloLens 2, which required them to turn their heads to see all possible paper positions on the table. To ensure that all sheets were comfortably reachable, the experimenter could adjust distances between papers during the tutorial so that even the farthest sheet remained within arm's reach. This setting was kept constant during the experiment.

The repeatability of each experiment was ensured by fixing the seeds on the random number generators. Moreover, potential biases were avoided by balancing out the color of the outline on the paper, ensuring that both colors occurred equally often within and between the left- and right-hand sides of the user.

The stamping task allowed users to interact with a physical surface that aligned with the virtual one, providing direct haptic feedback. This design choice was informed by insights from pilot experiments and previous work within our research group, highlighting the importance of tactile interaction [37].

To enhance engagement, a non-directional "bling" sound, similar to a phone notification, signaled the appearance of each new paper, while a chime rewarded users when they correctly stamped a sheet. Additionally, if a user could not find the paper within ten seconds, the system automatically advanced, causing the current sheet to disappear and the next one to appear.

Data Processing

For comprehensive behavioral analysis, accurate hand movement, gaze, and body position tracking were required. To address temporal and spatial misalignment, sensor noise, and data loss, the recorded data underwent extensive processing outlined in this section.

Temporal misalignment between the external tracking and the HoloLens 2 was corrected in two steps. During the recording, the hand poses were streamed to the HoloLens. An initial time offset estimate was obtained from the difference in sender and receiver timestamps. However, this still included network and processing delays and an error introduced by the difference in sampling rate (40 Hz vs 30 Hz). The time offset was therefore refined by maximizing cross-correlation of the normalized hand positions from both sources. Across all recordings, the cross-correlation of hand poses was on average 0.96, with external tracking data arriving with an average delay of 92 ms.

To address tracking losses and noise from either source, hand-tracking data from both the HoloLens 2 and external system were fused using a Kalman filter with a standard constant acceleration motion model. The two coordinate systems were initially aligned using a checkerboard calibration, but residual spatial errors were addressed by estimating a 3D offset within the filter. To avoid introducing temporal delays, Kalman smoothing was applied with a backward pass over the time series to achieve zero-phase filtering [38]. The filter parameters were tuned based on expected hand motion characteristics and system accuracies (Table 1). Innovation gating was used to reject outlier measurements that fell outside the 90% confidence region based on the calculated Mahalanobis distance.

Table . Kalman filter parameters.

Parameter	Value	Description
Process noise (acceleration)	3 m/s²	Expected SD of hand acceleration
Process noise (spatial offset)	0.0001 m/s	Random walk SD for coordinate alignment
External tracking measurement noise	0.02 m	SD of external tracking poses
HoloLens measurement noise	0.01 m	SD of HoloLens poses
Innovation gate threshold	90%	Confidence region for outlier rejection

Gaze data was recorded in two ways: On one hand, the gaze direction and origin were recorded, from which the construction of a gaze ray was possible. On the other hand, the intersections of the gaze ray with objects in the scene, for example, “Paper, Table, Ink Pad,” were computed at each update, and the target of the gaze was recorded. Given the reported variance of 1.5° to 3° in HoloLens 2 gaze directions [39], temporal filtering was applied to both time series. For the gaze target-object filtering, only Papers and Ink Pads were considered as targets of interest. A median filter over seven samples (spanning 0.175 s at 40 Hz) was applied to the time series. As a result, a target was considered fixated when first intersected, if subsequently the gaze remained on it for at least 50% of the sliding window duration (~0.085 s). The gaze origin and gaze direction vectors were filtered with a median filter with a 0.35 s window.

Head poses were obtained directly from the HoloLens 2 without filtering, as it exhibited little noise due to the device’s built-in robust localization system.

The torso rotation was reconstructed from the tracked chest, left, and right shoulder joints of the external tracking data. Outliers were removed based on sanity checks, asserting that the head position was always between the two tracked shoulder joints.

Metric Extraction and Data Visualization

To validate the system’s capability to measure common neglect symptoms, we focused on metrics that evaluate search behavior and attentional shifts therein. Our analysis incorporated spatial orientation metrics during the search phase (from paper appearance until target identification): average gaze direction, head rotation, and torso rotation, all computed relative to the table coordinate system using filtered sensor data and estimated forward vectors from the processed data.

For eye–hand coordination analysis, we examined the temporal relationship between gaze anchoring and motion onset. While gaze timing could be directly extracted from the gaze–target time series, determining precise motion onset required a more sophisticated approach.

An initial estimate, given by the recorded time the hand left the ink pad, was refined by identifying the optimal velocity minimum from candidate minima within the movement window. The selection was based on a multifactor scoring system that weighted (1) spatial proximity to the ink pad position, (2) hand velocity magnitude, (3) directional alignment with the target movement vector, and (4) temporal consistency with stamping events. The candidate with the lowest combined score was selected as the movement onset point.

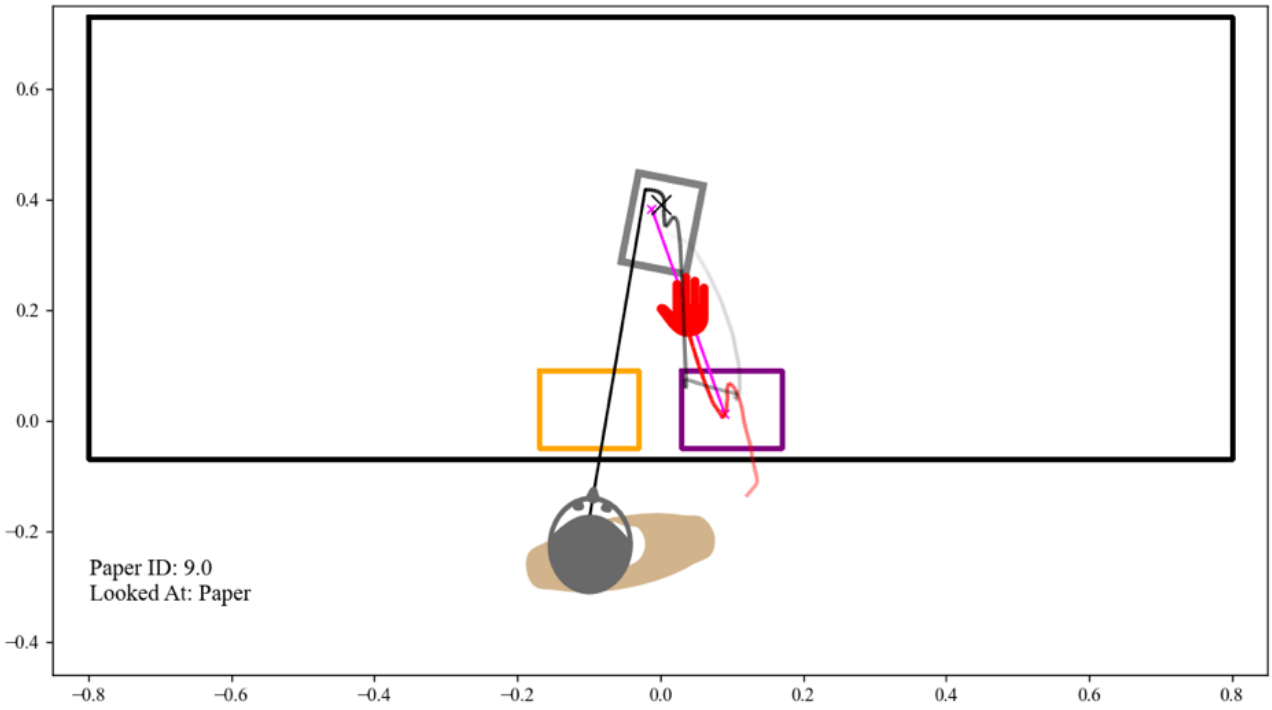
Despite this sophisticated approach, various technical and environmental factors could still affect motion onset detection accuracy, creating the need for a systematic data quality review. To address this, we implemented an automated outlier detection system that flagged an episode for review in case the duration of the motion exceeded 1.5 times the interquartile range (IQR) beyond the upper or lower quartile for individual user measurements (accrued metric over all episodes for that user).

Other metrics, such as the time it took to find a paper after its first appearance, were also processed by this system to detect anomalies resulting from technical issues such as occasional failures to register fixations or interactions.

To facilitate the review of these episodes and, in a broader sense, enable the qualitative analysis of the recorded interactions, we

implemented a digital twin: an interactive 2D visualization tool reconstructing user interactions (Figure 3). This digital twin provided an animated top-down view of the simulated environment, displaying the users' head, torso, and hand poses, gaze directions, and hand movement data. This allowed for a visual review of the flagged episodes to determine if the anomalies resulted from technical errors rather than natural human behavior. Any episodes where the visualization revealed tracking loss during critical phases, fixation or interaction registration failures, or other clear technical errors that compromised the affected metrics were marked as outliers. These episodes were excluded from the subsequent analysis.

Figure 3. Snapshot of digital twin replay of example data from our dataset. A black X marks the current fixation point. The magenta line connects the current motion start and end point. The torso with estimated orientation is shown in brown, together with the tracked positions of the hand and head. A red path shows the hand's position history, and a black path shows the user's gaze target over the past 3 seconds, both fading with time.



The digital twin tool abstracted sensitive information, enabling secure and collaborative qualitative analysis, which both clinicians and researchers found valuable for reviewing patient behavior. Its visualization of extracted metrics provided an intuitive way to replay and contextualize gaze, motion, and compensatory strategies, making it a practical aid for identifying patterns and generating insights to guide research and rehabilitation approaches.

Tracking Accuracy Evaluation

To evaluate the hand-tracking accuracy of the HoloLens 2, the ZED 2i body tracking, and the fused hand trajectories, we compared their 3D hand positions against ground-truth data recorded at 200 Hz with a VICON motion capture system.

Participants were equipped with infrared markers placed on both shoulders, the chest, and the index knuckle of each hand. Temporal alignment of the data was achieved by maximizing cross-correlation across speed and acceleration of hand movements. All signals were resampled to a common temporal grid. The spatial alignment was initialized by finding the rigid transformation that aligned the tracked shoulder and neck positions between the systems. It was then refined via an Iterative Closest Point (ICP)-like approach, maximizing global alignment of the hand positions between the Mocap system and the filtered hand positions.

Study Design

A proof-of-concept study was conducted with patients experiencing neglect symptoms after an RBL. Each patient played three rounds of the "game," where one round consisted of one paper appearing at each of the 21 locations, for a total of 63 episodes. The first round of the game was only played with the right hand; the subsequent two rounds presented the patients with papers containing left- and right-hand outlines, requiring patients to select the correct hand for the interaction. The order in which the papers appeared, as well as their color and hand outline, was randomized across games. The randomization was balanced to obtain two right-handed interactions per paper location for a total of 42 episodes (21 episodes in the first game, 12 in the second game, and 9 in the last game). The color randomization was equally balanced, resulting in a total of 33 episodes showing purple papers and 30 episodes with orange ones.

Before the experiment was started, the participants were asked to calibrate the eye tracker of the HoloLens 2 using Microsoft's calibration tool. To complete the calibration, a Vuforia marker [40] was used to align the holographic table with the physical table.

The patients were subsequently presented with a tutorial phase, during which no data were recorded, and the sheets did not disappear. The positions of the sheets in the tutorial were the

same as in the actual game, but the sequence and colors were different, to avoid learning effects. During this stage, the experimenter assisted patients in locating the sheets by referring to the external application monitor, which displayed the current sheet's position.

Once a participant had independently located and stamped six consecutive papers, the tutorial was terminated, and the actual assessment began. They were informed that the time was being recorded and instructed to perform the stamping task as fast and as accurately as possible.

While it was initially intended to also consider left-handed motions, we soon found that many of the patients had impairments to the contralesional limbs, which could not be attributed to USN. In the results, we therefore only considered the motion data from the right-handed interactions. For perception-only metrics, like the time until a paper was first looked at, all episodes were included in the analysis.

Ethical Considerations

The study was conducted in accordance with the latest version of the Declaration of Helsinki and received approval by the Ethics Committee of the Canton of Lucerne (ID 2017 - 02195). Before participation, all participants were informed about the content of the study, the steps and procedure of the experiment, and the goal of the study. Participants provided written informed consent and were explicitly informed of their right to revoke their consent at any time throughout the study or later. All data were anonymized to protect participants' privacy and confidentiality. Participants did not receive any compensation for their participation.

Usability

At the end of the experiment, the participants completed a usability questionnaire consisting of the following statements, each rated on a 5-point Likert scale where 1 was annotated with "strongly disagree" and 5 with "strongly agree (translated from German):

- Interacting with the elements (eg, paper, stamp) was comfortable
- I intuitively understood how to interact with the elements (paper, stamp).
- The game instructions were clear and easy to understand.
- The game maintained my attention from beginning to end.
- - I enjoyed playing the game.
- The challenges in the game were appropriate and well-balanced.
- I felt overwhelmed while playing the game.
- The game had a good pace—neither too slow nor too fast.
- I felt stressed by the game.

Results

Patient Characteristics

The study was conducted with seven right-handed patients (3 women and 4 men), between 49 and 83 years old (median 75 [58-78] y), who had an RBL. The participants were diagnosed with mild to moderate USN based on clinically relevant scores in a battery of tests (Table 2) consisting of mean gaze position (MGP) during free visual exploration (FVE) [18], bell's cancelation tests [13], and behavioral analysis using the Catherine Bergego Scale score [4,17]. While interactions with both hands were recorded, our analysis focused on the dominant and ipsilesional right hand. Only patients with little to no impairment of their right hand and arm movement were selected, which was assessed using the LIMOS score [41]. Due to time constraints, two patients (P1 and P5) completed only two out of three recordings during the experiment duration. For patient P5, this was anticipated, and they were asked to play the game with their right hand twice.

Table . USN group individual test scores on relevant clinical tests performed by clinicians prior to the study.

Patient ID	Gender	Age (y)	CBS[a]	MGP FVE[b]	CoC[c]	LIMOS[d] right hand and arm movement
P1	Female	83	4	3.92	0.04	4/5
P2	Male	49	1	1.45	0.02	5/5
P3	Male	79	0	0.43	0.15	4/5
P4	Female	77	7	3.25	0.37	5/5
P5	Male	67	5	1.33	0.09	5/5
P6	Female	74	12	2.52	0.11	5/5
P7	Male	49	0	2.00	0.09	4/5

[a]CBS: Catherine Bergego Scale (0% - 30%, higher=more neglect).

[b]MGP FVE: Mean gaze position during free exploration (cut-off>1.33°, rightward=positive).

[c]CoC: Center of cancellation (cutoff>0.081, left neglect=positive).

[d]LIMOS: 4=slowed and 5=independent.

The control group consisted of eight healthy individuals (1 female and 7 males) between 20 and 60 years old (median 31 [23-51] y). While none of the participants in the USN group had previous experience with AR devices, four out of the eight participants in the control group had previously used the technology.

Accuracy and Availability of Hand Tracking and Torso Rotation

Hand-tracking accuracy was evaluated for the HoloLens 2 and ZED 2i tracking system individually, as well as for their fused and filtered data, using the VICON Mocap system as ground truth. For evaluation, three male users independent from the clinical study participants were chosen, and measurements were obtained from gameplay with either hand, resulting in a total of 6 recordings (3 left hand and 3 right hand). We report the tracking accuracy during the relevant dynamic reaching motion only. The results in Table 3 show that the fused approach and the HoloLens 2 standalone achieved similar hand-tracking accuracies, with root mean square errors (RMSE) of 3.27 cm and 3.54 cm, respectively. The ZED 2i data contained large outliers (52.5 cm at the 95th percentile), which increased the RMSE to 37.1 cm. The median absolute error (AE) was also larger at 5.51 cm. The HoloLens 2 had frequent tracking loss, resulting in 74.3% availability. The ZED 2i and the fusion of the two sources achieved much better rates of 97.2% and 97.6%, respectively.

Table . Tracking accuracy results versus VICON data.

Tracking method	RMSE[a] (cm)	Median AE[b] (cm)	95th percentile AE[b] (cm)	Availability[c]
HoloLens 2	3.54	2.71	6.30	74.3%
ZED 2i	37.1	5.51	52.5	97.2%
Fusion	3.27	2.24	5.80	97.6%

[a]RMSE represents the root mean square average trajectory error during the reaching motion.

[b]Median AE and 95th percentile AE show the distribution of absolute errors across all samples.

[c]Availability indicates the percentage of frames with valid tracking data.

The torso rotation extracted from the ZED2i body tracking was compared to that obtained from the VICON system and evaluated during the search phase. The median signed error was 0.26°, with a median AE of 1.45° and 97.1% availability (Table 4).

Table . Torso rotation accuracy and availability against VICON data during free visual exploration[a].

Metric	Value
Median angle error (signed)	0.26°
Median absolute angle error	1.45°
95th percentile absolute angle error	4.41°
Availability	97.1%

[a]Availability indicates the percentage of frames with valid tracking data.

Omissions

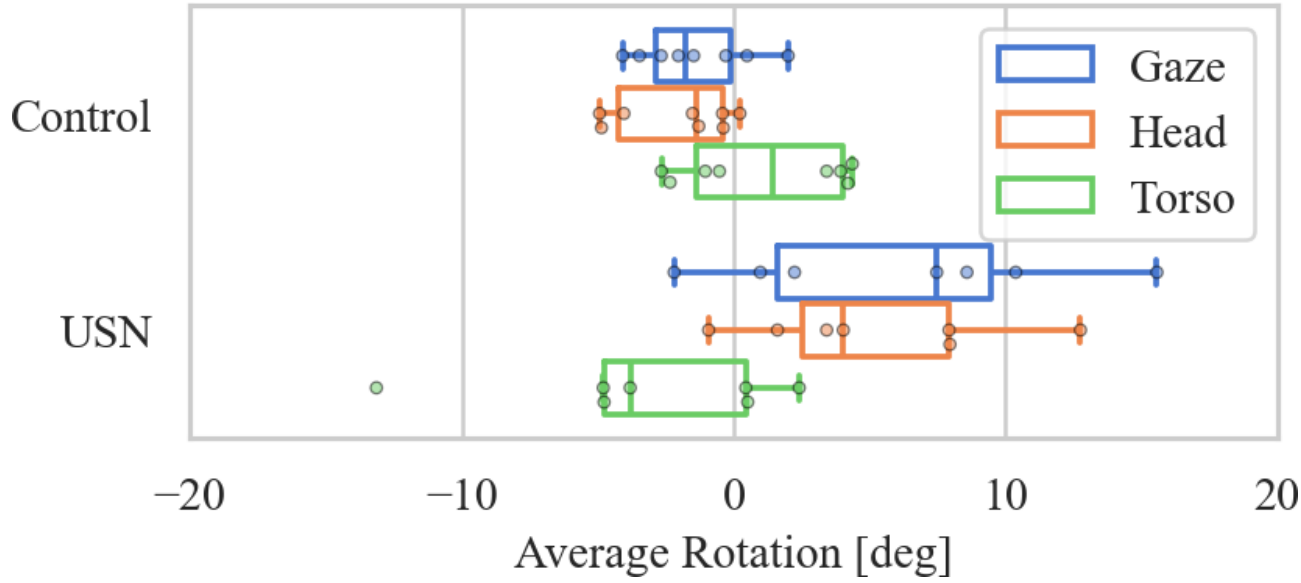
While patients received assistance and cues to search the entire space during the tutorial, no cues were provided during the actual experiment to avoid influencing search behavior. Papers that remained unfound after 10 s automatically disappeared, and the next paper appeared to maintain game progression. This time limit resulted in some omissions in the USN group: 9.3% (15/161 episodes) of papers on the left side and 3.3% (5/151 episodes) on the right side were not found within the time limit. The control group showed no omissions, nor were any omissions recorded for centrally positioned papers in either group.

Biases During Search Behavior

In line with previous research [8,9], the visual exploration of patients with USN was hypothesized to be biased toward the ipsilesional side. Using the proposed system, we analyzed the gaze direction during the search phase, as well as torso and head rotation. In the following, all statistical tests were performed using a Wilcoxon signed-rank test (due to the small number of participants) for a zero-mean hypothesis with a two-sided alternative. The level of significance was chosen at α=.05.

The results in Figure 4 show that the USN cohort exhibited a significant ipsilesional shift in the average gaze direction during FVE (median 7.46 [1.61-9.48] deg, *P*=.05). Similarly, a significant rightward trend of their average head rotation was observed (median 4.03 [2.52-7.94] deg, *P*=.03). A trend of an average torso rotation toward the contralesional side in the USN group did not achieve statistical significance (median −3.81 [−4.78 to 0.48] deg, *P*=.22) . In the control group, the average head orientation was shifted slightly, but significantly, toward the left during the search phase (median 1.40 [−4.25 to −0.04] deg , *P*=.02). Aside from that, the zero-mean hypothesis could not be rejected for the average gaze direction (median −1.75 [−2.87 to -0.10] deg, *P*=.11) and the average torso rotation (median 1.44 [−1.37 to 4.00] deg, *P*=.31) in the control group.

Figure 4. Boxplots and data points for average orientation of gaze, head, and torso relative to the table in front of the users. Boxes represent the IQR, the solid lines indicate the median, the whiskers indicate the minimum and maximum values that are not considered outliers.

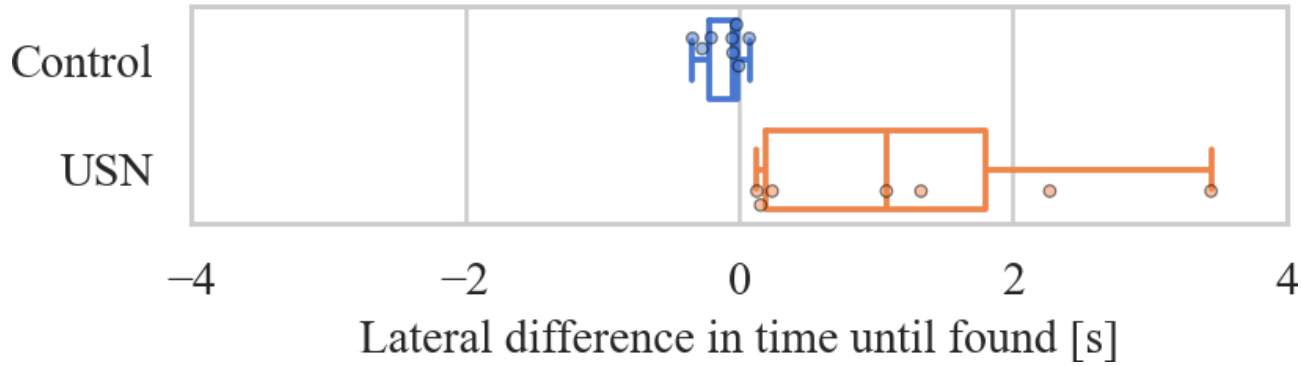


To investigate the effect of biased search behavior on the time it took users to find sheets of paper placed on the left and right sides of the table, we conducted a pairwise comparison. Each sheet on the left (within the two leftmost radial sectors) was matched with its mirrored counterpart on the right, and for each, we calculated the average time it took a user to find it across multiple trials. We then compared the average time for each paper on the left to the average time for its counterpart on the right. For example, if a user took an average of 8 s to find a specific sheet on the left and an average of 5 s to find its counterpart on the right, the difference for that pair would be +3 s.

This difference was computed for each pair and then averaged across all pairs per user to yield an overall measure of lateral difference. In the control group, out of the total 504 recorded episodes, 64 were flagged and removed as outliers (12.7%), whereas 50 of the total 399 episodes were removed in the USN group (12.5%). Any pairs with no data on either side were excluded from the analysis.

The results (Figure 5) indicate that the USN group took significantly longer to find papers on the left side compared to the right (median difference 1.08 s [0.20-1.80], *P*=.02), whereas no significant difference in the control group was observed.

Figure 5. Pairwise differences of the time until each paper was found on the left and right side. A positive difference indicates that the papers on the right were found sooner. Box represents the IQR, the solid line indicates the median, the whiskers indicate the minimum and maximum values that are not considered outliers. Control: n=8, median −0.05 s [−0.22 to −0.01], *P*=.08; USN: n=7; median 1.08 s [0.20-1.80], *P*=.02.



Differences in Timing of Gaze Anchoring

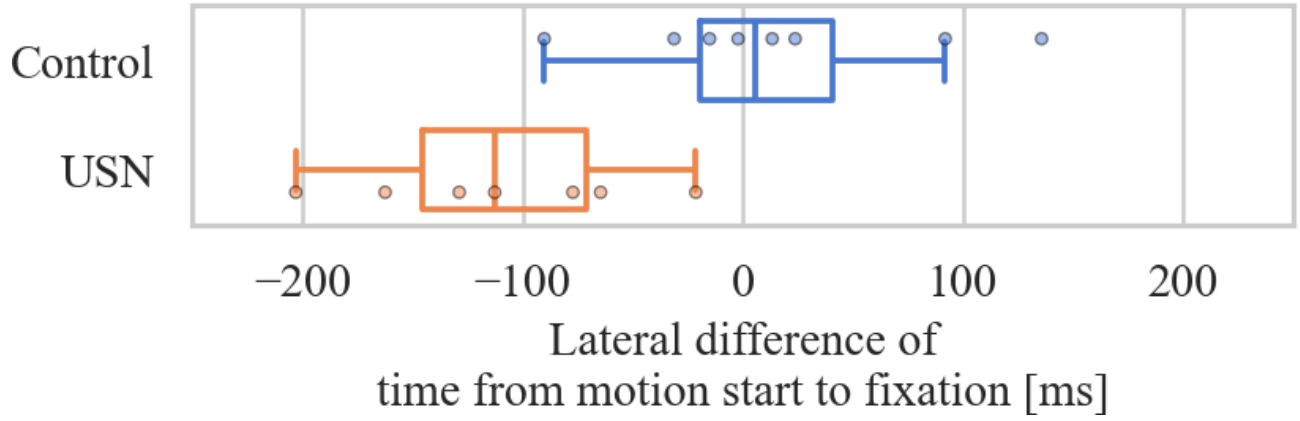
As an important aspect of eye–hand coordination, we investigate gaze anchoring behavior during targeted motion. A typical episode involved the following steps: (1) the participants were located and fixated on the paper; (2) gaze was shifted to the ink pad while the participant selected a color with their hand; (3) hand motion was initiated away from the ink pad toward the remembered location of the paper; and (4) gaze anchoring was shifted back to the paper to guide the ongoing motion.

Eye–hand coordination is a complex field of research. One aspect that is often considered is the timing of fixations with respect to intent and concrete actions [37]. In patients with USN, disrupted spatial attention may particularly affect this temporal coordination between gaze and movement. In the following, we therefore analyzed lateral time differences between (3) and (4). This was done using a pairwise comparison, similar to the previous section. For each side, we calculated the average time between the onset of hand movement and the fixation of the target paper, comparing these times between left- and right-hand sides. For example, if a user fixated on a sheet 100 ms after

initiating hand movement toward it on the left side and 300 ms on the right, the difference for this pair would be −200 ms. The outlier review resulted in 52 of the 336 episodes with right-handed interaction data (15.5%) being rejected in the control cohort and 47 out of 246 (19.1%) in the USN group.

The results (Figure 6) show that the USN group fixated on the target of their motion significantly earlier for targets on their left compared to their right (median difference of 112 ms [-146 to -71], *P*=.02). In contrast, the control showed no significant lateral difference (*P*=.74).

Figure 6. Pairwise difference between the sheets on the left and on the right side and of the time between motion onset and fixation of the motion target. A negative difference indicates that sheets on the left were fixated earlier. The box represents the IQR, the solid line indicates the median, and the whiskers indicate the minimum and maximum values that are not considered outliers. Control: n=8, median 5 ms [-19 to 41], *P*=.74; USN: n=7, median −112 ms [-146 to -71], *P*=.02.



Given that gaze anchoring has been shown to shift in time with longer duration movements [42], we examined whether the observed laterality in gaze timing could be explained by corresponding differences in movement durations. To assess this, we analyzed the relationship between lateral differences in gaze timing and lateral differences in the duration of reaching motions across participants. Pearson correlation analysis revealed a moderate correlation, not achieving statistical significance (control: *r*=.20, *P*=.63; USN: *r*=.67, *P*=.10). Additionally, we examined whether patients with USN showed systematic differences in motion duration between left and right sides. Despite a mean difference of −80 ms, the median difference was effectively zero (0.8 [-204 to 63] ms), and a zero mean hypothesis could not be rejected (*P*=.58).

Usability

Table 5 shows the median ratings for all usability questions. Patients rated the game as comfortable (median 4) and intuitive (median 5), with high scores for clear instructions (median 5), enjoyable gameplay (median 5), and balanced challenges (median 4). Participants generally felt the game was engaging (median 4) and appropriately paced (median 4), with minimal reports of stress (median 1) or overwhelm (median 1).

Table . Overview of the usability study questions and answers.

Question	Median
Interacting with the elements (eg, paper, stamp) was comfortable.	4
I intuitively understood how to interact with the elements (paper, stamp).	5
The game instructions were clear and easy to understand.	5
The game maintained my attention from beginning to end.	4
I enjoyed playing the game.	5
The challenges in the game were appropriate and well-balanced.	4
I felt overwhelmed while playing the game.	1
I was motivated to continue playing.	4
The game had a good pace—neither too slow nor too fast.	4
I felt stressed by the game.	1

Discussion

Accuracy and Availability of Hand Tracking and Torso Rotation

The HoloLens standalone method clearly demonstrates the tracking loss issue described in the Introduction. The data from the ZED 2i, while achieving excellent availability, suffered from both higher systematic errors, likely due to coordinate system misalignment, and significant noise, resulting in poor overall accuracy. Our fusion approach combined advantages of both sources, achieving better tracking accuracy than the HoloLens 2 while maintaining near-complete availability.

Torso angle tracking achieved high availability with no systematic bias, as evidenced by the small median signed angle error (0.26°). However, the median absolute angle error of 1.45° is relatively large compared to the clinical effects found during the search behavior phase, limiting the confidence in these results. This could possibly be improved in the future by ensuring better alignment between the external tracking system and the HoloLens 2 coordinate system.

Biased Search Behavior in Patients With USN

The analysis of the MGP during the search phase revealed, as expected, an ipsilesional shift in the patients with USN with respect to the table center. This was consistent with previously recorded behavior of patients with USN [43-45]. The shifted MGP also explains the prolonged time until papers on the contralesional side were found.

A significant feature of our method compared to traditional eye tracking during FVE is that it allows for free rotation of the head and torso. Analyzing the search behavior with this approach revealed that patients exhibited a bias of their head rotation toward the ipsilesional side, aligning with their gaze direction. Moreover, some patients rotated their trunks slightly toward the contralesional side, although this was not a behavior that achieved significance among the USN group. Nonetheless, this prompts further investigation into the observed behavior, as trunk orientation is known to define the reference frame for visual neglect. Hence, a rotation toward the contralesional side could have reduced the attentional bias to a certain degree in those patients, as reported by [19,20,46]. In general, these findings emphasize the importance of measuring torso rotation for a comprehensive assessment of USN symptoms.

The control group exhibited a slight leftward bias of their head rotation relative to the table, which can be explained by the intrinsic left-to-right scanning tendency, sometimes associated with reading direction, and right-hemispheric dominance for attention prevalent among healthy individuals [47,48].

Lateral Differences in Timing of Gaze Anchoring

The observed difference in gaze anchoring timing between the left and right sides encourages the hypothesis that neglect may impact eye–hand coordination in the USN group, though this remains a preliminary finding. In both groups, users generally initiated the motion from the ink pad to the paper "blindly," relying initially on the remembered position of the paper. However, in the USN group, the results indicate that gaze anchoring occurs earlier in the motion for targets on the left compared to the right.

Several factors may contribute to this asymmetry. First, the USN group might have put less trust in their memorized target positions on the left side, prompting them to fixate on those targets sooner. Supporting this theory, previous works have reported increased localization errors on the contralesional side [49]. According to [50], the visuospatial working memory was negatively affected after patients shifted their attention toward stimuli ipsilesional to the memorized location. In our experiment setup, such a stimulus was present, as the stamping task involved an attentional shift back to the centrally positioned ink pads, requiring patients to memorize a contralesional location. Additionally, reaching with the right hand toward a target on the left may also have induced an intermediate attentional shift toward the ipsilesional side, reinforcing this effect.

Second, differences in movement duration could partially account for the gaze anchoring asymmetry [42]. We observed a moderate correlation between lateral differences in gaze timing and movement in the USN group (r=.67, P=.10). However, the movement duration differences showed high variability and no consistent lateral pattern. This suggests that those differences cannot fully explain the gaze timing asymmetry. Given the small sample size, further investigation with a larger cohort is needed to clarify this potential contribution.

Usability

The positive feedback across several aspects of the framework and game suggests that it was accessible and motivating for users, supporting the feasibility of using it for more extensive clinical trials and during rehabilitation with patients with USN.

Limitations

First, the lack of a stroke-only control group (without USN) limits our ability to confidently attribute the observed effects specifically to USN. Additionally, substantial demographic disparities existed between groups, which are known to influence motor performance and limit the comparability of motor-related metrics between groups. Future studies should include a stroke control group without USN, matched on demographics, to better isolate USN-specific effects.

Second, both groups had relatively small sample sizes, with only seven patients in the USN group whose neglect severity varied from mild to moderate. This might have inflated the large effect size (Cohen d≈2.3) of the lateral difference in gaze-anchoring timing in the USN group. Replication with a larger cohort may be necessary to validate these findings.

Third, measurement precision of torso rotation was comparatively low and limits the confidence in the associated results. Fourth, to obtain the tracking accuracy results, the VICON data had to be spatially and temporally aligned to the HoloLens in postprocessing. Therefore, these results reflect tracking accuracy within the HoloLens' coordinate system rather than absolute world coordinates.

Despite these limitations, the results align with the hypothesized behavior of patients with USN, strengthening our confidence in the validity of the results.

Conclusions

In this work, we developed and validated a comprehensive behavioral tracking system for patients with stroke and USN that can capture eye, body, and hand movements during task execution in a mixed reality environment. By integrating a head-mounted AR display with an external body tracking system, we obtained naturalistic, multimodal interaction data in patients with stroke and USN. The system combines a game-like task with automated metric extraction and data visualization, offering a scalable tool for quantitative and qualitative behavioral assessment.

Importantly, the system demonstrated a significant improvement in hand-tracking accuracy and availability over the standalone hand tracking of the HoloLens 2 or ZED 2i, with a fused hand-tracking error of 3.27 cm RMSE at 2.4% tracking loss.

In a proof-of-concept study with seven RBL patients with mild to moderate USN, the framework captured behavior consistent with known USN symptoms, such as ipsilesional gaze biases and delayed contralesional target detection.

The naturalistic task design combined with our system's multimodal data capture revealed new insights into USN visuomotor behavior. In particular, we discovered a pattern where patients directed their gaze toward movement targets earlier when those targets appeared on their contralesional side compared to their ipsilesional side.

This framework opens several promising research and clinical avenues. Larger patient cohorts and appropriate control groups could provide more robust quantitative insights into USN-related visuomotor deficits. Future investigations of nondominant hand use and bimanual coordination, and more extensive movement kinematic analysis, could reveal additional layers of motor-cognitive interaction in stroke recovery. Ultimately, this framework lays the groundwork for a truly comprehensive assessment of how USN manifests across the entire visuomotor cascade in a naturalistic setting.

Acknowledgments

We would like to warmly thank all the patients and volunteers who participated in this study.

Generative AI Disclosure: All content of this work was conceived and written by the authors. Large language models were used as editing tools to refine the manuscript's language, specifically: (1) GPT-4o (2024) and Claude Sonnet 4.0 (2025) were used to review the text for grammatical errors and to suggest alternative phrasings for improved clarity and readability of author-written content. (2) GitHub Copilot (2023) was used as a coding assistant to suggest code completions during software development. No AI-generated content was used verbatim. All AI suggestions were critically reviewed, modified as needed, and approved by the authors, who retain full ownership and responsibility for the manuscript's content and accuracy. We would like to warmly thank all the patients and volunteers who participated in this study.

This work was supported by the ETH RESC - Suva Funding Programme.

Data Availability

Public archiving of the data is not permitted under our ethics approval. In accordance with the Swiss Human Research Act, readers who wish to access the data and study materials should contact the corresponding author and complete a formal data-sharing agreement.

Authors' Contributions

JB and SK developed the app and implemented data processing. JB, DC, and QL conceived and designed the study. JB, DC, and TN recruited the patients and collected the data. JB, SK, DC, and QL analyzed and interpreted the data. JB, SK, DC, and QL drafted the report. All authors contributed to reviewing it; all authors read and approved the final manuscript.

Conflicts of Interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript.

References

1. Feigin VL, Brainin M, Norrving B, et al. World Stroke Organization (WSO): global stroke fact sheet 2022. Int J Stroke 2022 Jan;17(1):18-29. [doi: 10.1177/17474930211065917] [Medline: 34986727]
2. Ringman JM, Saver JL, Woolson RF, Clarke WR, Adams HP. Frequency, risk factors, anatomy, and course of unilateral neglect in an acute stroke cohort. Neurology (ECronicon) 2004 Aug 10;63(3):468-474. [doi: 10.1212/01.wnl.0000133011.10689.ce] [Medline: 15304577]
3. Excellence stroke rehabilitation in adults. National Institute for Health and Care. URL: https://www.nice.org.uk/guidance/ng236/chapter/Recommendations [accessed 2025-10-18]
4. Azouvi P, Olivier S, de Montety G, Samuel C, Louis-Dreyfus A, Tesio L. Behavioral assessment of unilateral neglect: study of the psychometric properties of the Catherine Bergego Scale. Arch Phys Med Rehabil 2003 Jan;84(1):51-57. [doi: 10.1053/apmr.2003.50062] [Medline: 12589620]
5. Jehkonen M, Ahonen JP, Dastidar P, et al. Visual neglect as a predictor of functional outcome one year after stroke. Acta Neurol Scand 2000 Mar;101(3):195-201. [doi: 10.1034/j.1600-0404.2000.101003195.x] [Medline: 10705943]
6. Katz N, Hartman-Maeir A, Ring H, Soroker N. Functional disability and rehabilitation outcome in right hemisphere damaged patients with and without unilateral spatial neglect. Arch Phys Med Rehabil 1999 Apr;80(4):379-384. [doi: 10.1016/s0003-9993(99)90273-3] [Medline: 10206598]

7. Chen P, Hreha K, Kong Y, Barrett AM. Impact of spatial neglect on stroke rehabilitation: evidence from the setting of an inpatient rehabilitation facility. Arch Phys Med Rehabil 2015 Aug;96(8):1458-1466. [doi: 10.1016/j.apmr.2015.03.019] [Medline: 25862254]
8. Demeyere N, Gillebert CR. Ego- and allocentric visuospatial neglect: dissociations, prevalence, and laterality in acute stroke. Neuropsychology 2019 May;33(4):490-498. [doi: 10.1037/neu0000527] [Medline: 30896236]
9. Fruhmann-Berger M, Karnath HO. Spontaneous eye and head position in patients with spatial neglect. J Neurol 2005 Oct;252(10):1194-1200. [doi: 10.1007/s00415-005-0831-y] [Medline: 15895307]
10. Husain M, Rorden C. Non-spatially lateralized mechanisms in hemispatial neglect. Nat Rev Neurosci 2003 Jan;4(1):26-36. [doi: 10.1038/nrn1005] [Medline: 12511859]
11. Karnath HO, Rorden C. The anatomy of spatial neglect. Neuropsychologia 2012 May;50(6):1010-1017. [doi: 10.1016/j.neuropsychologia.2011.06.027] [Medline: 21756924]
12. Williams LJ, Kernot J, Hillier SL, Loetscher T. Spatial neglect subtypes, definitions and assessment tools: a scoping review. Front Neurol 2021;12:742365. [doi: 10.3389/fneur.2021.742365] [Medline: 34899565]
13. Gauthier L, Dehaut F, Joanette Y, et al. The bells test: a quantitative and qualitative test for visual neglect. Int J Clin Neuropsychol 1989;11(2):49-54 [FREE Full text]
14. Schenkenberg T, Bradford DC, Ajax ET. Line bisection and unilateral visual neglect in patients with neurologic impairment. Neurology (ECronicon) 1980 May;30(5):509-517. [doi: 10.1212/wnl.30.5.509] [Medline: 7189256]
15. Lindell AB, Jalas MJ, Tenovuo O, Brunila T, Voeten MJM, Hämäläinen H. Clinical assessment of hemispatial neglect: evaluation of different measures and dimensions. Clin Neuropsychol 2007 May;21(3):479-497. [doi: 10.1080/13854040600630061] [Medline: 17455032]
16. Azouvi P, Samuel C, Louis-Dreyfus A, et al. Sensitivity of clinical and behavioural tests of spatial neglect after right hemisphere stroke. J Neurol Neurosurg Psychiatry 2002 Aug;73(2):160-166. [doi: 10.1136/jnnp.73.2.160] [Medline: 12122175]
17. Bergego C, Azouvi P, Samuel C, et al. Validation d'une échelle d'évaluation fonctionnelle de l'héminégligence dans la vie quotidienne: l'échelle CB [Article in French]. Annales de Réadaptation et de Médecine Physique 1995;38(4):183-189. [doi: 10.1016/0168-6054(96)89317-2]
18. Kaufmann BC, Cazzoli D, Pflugshaupt T, et al. Eyetracking during free visual exploration detects neglect more reliably than paper-pencil tests. Cortex 2020 Aug;129:223-235. [doi: 10.1016/j.cortex.2020.04.021] [Medline: 32512414]
19. Karnath HO, Schenkel P, Fischer B. Trunk orientation as the determining factor of the "contralateral" deficit in the neglect syndrome and as the physical anchor of the internal representation of body orientation in space. Brain (Bacau) 1991 Aug;114 (Pt 4)(4):1997-2014. [doi: 10.1093/brain/114.4.1997] [Medline: 1884189]
20. Karnath HO, Christ K, Hartje W. Decrease of contralateral neglect by neck muscle vibration and spatial orientation of trunk midline. Brain (Bacau) 1993 Apr;116 (Pt 2)(2):383-396. [doi: 10.1093/brain/116.2.383] [Medline: 8461972]
21. Rorden C, Li D, Karnath HO. Biased temporal order judgments in chronic neglect influenced by trunk position. Cortex 2018 Feb;99:273-280. [doi: 10.1016/j.cortex.2017.12.008] [Medline: 29306707]
22. Shida K, Amimoto K, Fukata K, Osaki S, Takahashi H, Makita S. The effect of trunk position on attentional disengagement in unilateral spatial neglect. Neurol Int 2022 Dec 13;14(4):1036-1045. [doi: 10.3390/neurolint14040083] [Medline: 36548188]
23. Jackson CPT, Miall RC, Balslev D. Spatially valid proprioceptive cues improve the detection of a visual stimulus. Exp Brain Res 2010 Aug;205(1):31-40. [doi: 10.1007/s00221-010-2330-3] [Medline: 20567807]
24. Balslev D, Odoj B. Distorted gaze direction input to attentional priority map in spatial neglect. Neuropsychologia 2019 Aug;131:119-128. [doi: 10.1016/j.neuropsychologia.2019.05.017] [Medline: 31128129]
25. Cavedoni S, Cipresso P, Mancuso V, Bruni F, Pedroli E. Virtual reality for the assessment and rehabilitation of neglect: where are we now? A 6-year review update. Virtual Real 2022;26(4):1663-1704. [doi: 10.1007/s10055-022-00648-0] [Medline: 35669614]
26. Arntz A, Weber F, Handgraaf M, et al. Technologies in home-based digital rehabilitation: scoping review. JMIR Rehabil Assist Technol 2023 Jul 27;10:e43615. [doi: 10.2196/43615] [Medline: 37253381]
27. Bakker MDJ, Boonstra N, Nijboer TCW, Holstege MS, Achterberg WP, Chavannes NH. The design choices for the development of an augmented reality game for people with visuospatial neglect. Clin eHealth 2020;3:82-88. [doi: 10.1016/j.ceh.2020.11.003]
28. Gorman C, Gustafsson L. The use of augmented reality for rehabilitation after stroke: a narrative review. Disabil Rehabil Assist Technol 2022 May;17(4):409-417. [doi: 10.1080/17483107.2020.1791264] [Medline: 32663112]
29. Held JPO, Yu K, Pyles C, et al. Augmented reality-based rehabilitation of gait impairments: case report. JMIR Mhealth Uhealth 2020 May 26;8(5):e17804. [doi: 10.2196/17804] [Medline: 32452815]
30. Monica R, Rizzini DL, Aleotti J. Adaptive complementary filter for hybrid inside-out outside-in HMD tracking with smooth transitions. IEEE Trans Vis Comput Graph 2025 Feb;31(2):1598-1612. [doi: 10.1109/TVCG.2024.3464738] [Medline: 39298310]
31. Gsaxner C, Li J, Pepe A, et al. The HoloLens in medicine: a systematic review and taxonomy. Med Image Anal 2023 Apr;85:102757. [doi: 10.1016/j.media.2023.102757] [Medline: 36706637]

32. Vávra P, Roman J, Zonča P, et al. Recent development of augmented reality in surgery: a review. J Healthc Eng 2017;2017:4574172. [doi: 10.1155/2017/4574172] [Medline: 29065604]
33. Palumbo A. Microsoft HoloLens 2 in medical and healthcare context: state of the art and future prospects. Sensors (Basel) 2022 Oct 11;22(20):7709. [doi: 10.3390/s22207709] [Medline: 36298059]
34. HoloLens 2—overview and technical details. Microsoft. URL: https://www.microsoft.com/de-de/hololens/hardware [accessed 2024-11-21]
35. MixedRealityToolkit (MRTK). GitHub/Microsoft. URL: https://github.com/microsoft/MixedRealityToolkit [accessed 2024-11-21]
36. ZED SDK. GitHub/stereolabs. URL: https://github.com/stereolabs/zed-sdk [accessed 2024-11-21]
37. Wolf J, Lohmeyer Q, Holz C, Meboldt M. Gaze comes in handy: predicting and preventing erroneous hand actions in AR-supported manual tasks. Presented at: 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); Oct 4-8, 2021; Bari, Italy p. 166-175. [doi: 10.1109/ISMAR52148.2021.00031]
38. Pykalman. GitHub/pykalman. URL: https://github.com/pykalman/pykalman [accessed 2024-11-21]
39. Eye tracking overview – mixed reality. Microsoft. URL: https://learn.microsoft.com/en-us/windows/mixed-reality/design/eye-tracking [accessed 2024-11-21]
40. Vuforia enterprise augmented reality (AR) software. PTC (DE). URL: https://www.ptc.com/de/products/vuforia [accessed 2024-11-21]
41. Ottiger B, Vanbellingen T, Gabriel C, et al. Validation of the new Lucerne ICF based Multidisciplinary Observation Scale (LIMOS) for stroke patients. PLoS ONE 2015;10(6):e0130925. [doi: 10.1371/journal.pone.0130925] [Medline: 26110769]
42. Lavoie EB, Valevicius AM, Boser QA, et al. Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. J Vis 2018 Jun 1;18(6):18. [doi: 10.1167/18.6.18] [Medline: 30029228]
43. Karnath HO, Niemeier M. Task-dependent differences in the exploratory behaviour of patients with spatial neglect. Neuropsychologia 2002;40(9):1577-1585. [doi: 10.1016/s0028-3932(02)00020-9] [Medline: 11985839]
44. Schindler I, Clavagnier S, Karnath HO, Derex L, Perenin MT. A common basis for visual and tactile exploration deficits in spatial neglect? Neuropsychologia 2006;44(8):1444-1451. [doi: 10.1016/j.neuropsychologia.2005.12.003] [Medline: 16499936]
45. Olk B, Harvey M. Characterizing exploration behavior in spatial neglect: omissions and repetitive search. Brain Res 2006 Nov 6;1118(1):106-115. [doi: 10.1016/j.brainres.2006.08.011] [Medline: 16979143]
46. Schindler I, Kerkhoff G. Head and trunk orientation modulate visual neglect. Neuroreport 1997 Aug 18;8(12):2681-2685. [doi: 10.1097/00001756-199708180-00009] [Medline: 9295100]
47. Nicholls MER, Roberts GR. Can free-viewing perceptual asymmetries be explained by scanning, pre-motor or attentional biases? Cortex 2002 Apr;38(2):113-136. [doi: 10.1016/s0010-9452(08)70645-2] [Medline: 12056684]
48. Shulman GL, Pope DLW, Astafiev SV, McAvoy MP, Snyder AZ, Corbetta M. Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. J Neurosci 2010 Mar 10;30(10):3640-3651. [doi: 10.1523/JNEUROSCI.4085-09.2010] [Medline: 20219998]
49. Cohen-Dallal H, Soroker N, Pertzov Y. Working memory in unilateral spatial neglect: evidence for impaired binding of object identity and object location. J Cogn Neurosci 2021 Jan;33(1):46-62. [doi: 10.1162/jocn_a_01631] [Medline: 32985947]
50. Vuilleumier P, Sergent C, Schwartz S, et al. Impaired perceptual memory of locations across gaze-shifts in patients with unilateral spatial neglect. J Cogn Neurosci 2007 Aug;19(8):1388-1406. [doi: 10.1162/jocn.2007.19.8.1388] [Medline: 17651010]

Abbreviations

AE: absolute error
AR: augmented reality
ATE: absolute trajectory error
FVE: free visual exploration
MGP: mean gaze position
Mocap: motion capture
RBL: right brain lesion
USN: unilateral spatial neglect
VR: virtual reality

Edited by T Leung, L Riedemann; submitted 07.01.25; peer-reviewed by B Dexheimer, CS Chapman, M Edwards; revised version received 11.10.25; accepted 13.10.25; published 03.11.25.

Please cite as:
Becker J, Ktistakis S, Meboldt M, Nyffeler T, Cazzoli D, Lohmeyer Q
Augmented Reality Framework to Measure and Analyze Eye–Hand Coordination in Stroke Patients with Unilateral Neglect: Proof-of-Concept Study
JMIR XR Spatial Comput 2025;2:e70985
URL: https://xr.jmir.org/2025/1/e70985
doi:10.2196/70985

Assessing the Feasibility of Using Apple Vision Pro While Performing Medical Precision Tasks: Controlled User Study

Hamraz Javaheri[1], MSc; Vitor Fortes Rey[1,2], Dr Ing; Paul Lukowicz[1,2], Dr rer nat; Gregor A Stavrou[3], Dr med, Dr habil; Jakob Karolus[1,2], Dr rer nat; Omid Ghamarnejad[3], Dr med

[1]Department of Embedded Intelligence, German Research Centre for Artificial Intelligence, Trippstadter Str. 122, Kaiserslautern, Germany

[2]Department of Computer Science, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Kaiserslautern, Germany

[3]Department of General, Visceral, and Oncological Surgery, Klinikum Saarbrücken, Saarbrücken, Germany

Corresponding Author:
Hamraz Javaheri, MSc
Department of Embedded Intelligence, German Research Centre for Artificial Intelligence, Trippstadter Str. 122, Kaiserslautern, Germany

Abstract

Background: The emergence of next-generation video-see-through head-mounted displays, such as the Apple Vision Pro (AVP), has generated considerable interest in the medical field. While preliminary studies highlight AVP's potential, no controlled study has rigorously assessed its usability for precision-based medical tasks requiring fine motor control and real-world perception.

Objective: This study aims to evaluate the feasibility of using AVP while performing real-world medical precision tasks.

Methods: To assess AVP's feasibility, we conducted a controlled user study with 20 health care professionals, who performed 3 different suturing techniques across 3 intervention conditions. Participants completed the same tasks using AVP, the Microsoft HoloLens 2 (MHL2), and a baseline (without a head-mounted display). A within-subject design was used, ensuring that each participant experienced all intervention groups. We used a mixed methods research approach, incorporating both quantitative metrics, including task completion time, suturing performance, system usability score, cognitive load, virtual reality sickness, and presence score, as well as qualitative insights gathered through interviews.

Results: Participants took significantly longer to complete the entire task using AVP (570.0, SD 192.0 s) compared with MHL2 (456.0, SD 120.0 s; $P<.001$) and baseline (472.0, SD 143.0 s; $P<.001$). The analysis on participants' average suture performance revealed no significant differences across interventions ($P=.76$). The total raw NASA Task Load Index score among participants was significantly higher for AVP (43.9, SD 15.9) compared with MHL2 (21.5, SD 13.8; $P<.001$) and baseline (19.1, SD 15.1; $P<.001$). The analysis of the presence questionnaire demonstrated a significantly higher presence score for MHL2 (115.0, SD 11.4) compared with AVP (93.7, SD 12.7; $P<.001$). The overall virtual reality sickness questionnaire score was significantly higher for AVP (66.9, SD 19.8) compared with MHL2 (41.1, SD 9.32; $P<.001$). Moreover, the calculated system usability score for MHL2 (72.7, SD 8.54) was significantly higher compared with AVP (50.3, SD 14.4; $P<.001$).

Conclusions: In conclusion, AVP has potential for non–time-sensitive medical applications or those that emphasize digital elements over real-world interaction. Its current usability limitations, particularly increased cognitive load and prolonged task execution times, suggest that further optimizations are necessary before widespread clinical adoption is feasible.

(JMIR XR Spatial Comput 2025;2:e73574) doi:10.2196/73574

KEYWORDS

Apple Vision Pro; HoloLens; extended reality; precision task; head-mounted display

Introduction

Mixed reality (MR) has been a transformative technology for several years, revolutionizing various industries and applications. As part of the broader spectrum of immersive technologies [1], MR bridges augmented reality (AR), which overlays digital content onto the real world, and virtual reality (VR), which provides fully immersive digital environments. With recent advancements in wearable technology and head-mounted displays (HMDs), MR has expanded into a wide range of daily activities and professional domains [2]. A significant development in this field is the rise of extended reality (XR) devices, which integrate both AR and VR capabilities, enabling seamless transitions between immersive and real-world experiences. This new generation of devices, such as the Apple Vision Pro (AVP) [3], has sparked considerable interest and is believed to be the future of HMDs in the medical domain, offering users immersive XR experiences through video-see-through (VST) technology [4-12]. Although there are studies that have examined XR applications across

various settings and domains, the choice of device and technology for specific task groups is often driven by the latest market trends rather than an informed assessment of their feasibility for the intended use case.

Egger et al [4,5] regarded AVP as a major step toward achieving the "ultimate display" for health care. They highlighted its potential to address challenges that previous MR devices, such as the Microsoft HoloLens (MHL) [13], encountered in terms of precision, reliability, usability, workflow integration, and user perception. Similarly, Masalkhi et al [6] postulated that Apple XR technology holds a wide range of possibilities in ophthalmology, including applications in surgical training, assistive devices, diagnosis, and education. Furthermore, Olexa et al [7] reported using AVP as a neurosurgical planning tool to visualize 3D models of patients. They noted that users found the 3D models to be highly realistic (Likert score of 4.5/5), the real-world view displayed through the headset to be natural (Likert score of 4.3/5), and experienced minimal eye strain or fatigue while using the device.

While these studies highlight the significant potential of AVP in the medical field and its ability to support the development of various beneficial applications, no controlled or experimental study has rigorously evaluated the usability of the device itself and its impact on user real-world performance and experience, independent of the application used. In particular, the suitability of AVP and its VST design remains unexplored in scenarios that involve wearing the headset while performing real-world, delicate tasks requiring fine motor control and precision, such as those encountered in surgical support and navigation systems. As immersive technologies continue to evolve, the extent to which the choice of device directly influences user performance remains unclear. Furthermore, while VST devices, such as AVP, are expected to enhance the accuracy and precision required for medical tasks by providing higher-quality visualization and more robust registration of digital objects, the usability of these technologies and the devices themselves for applications requiring real-world precision remains uncertain, as users rely on visual information through a video stream rather than direct visual perception. Consequently, the safety and feasibility of AVP in critical applications, such as in situ surgical navigational systems, need to be investigated. Additionally, its proclaimed superiority over existing HMDs in medical domains, which demand high levels of accuracy and dexterity, remains unproven. This gap presents an important opportunity for further research to substantiate the benefits and advantages of different XR approaches, including optical see-through (OST) and VST.

Many of the applications envisioned for AVP have already been achieved using MHL, an OST MR device series that has demonstrated a broad range of applications in medicine. These include patient data visualization [14], patient education [15], assistance and monitoring [16], preoperative diagnosis [17], anatomy learning [18,19], intervention training [20], image-guided interventions [21], in situ surgical navigation [22-26], and telemedicine [27]. Extensive use has demonstrated the effectiveness of MHL in these areas and provides a strong baseline for evaluating new XR devices, such as AVP, particularly for medical applications that rely on MR capabilities rather than pure VR and require visual perception of the real world.

This study addresses these gaps by evaluating the feasibility of AVP for performing medical precision tasks that require visual perception of the real world. The main goal of this study is to evaluate the feasibility of the devices and their underlying technology without any bias from specific applications. This work contributes to understanding the impact of device choice on user performance by examining both objective performance metrics and subjective measures of user experience. To achieve this, we designed a controlled user study involving 20 health care professionals, comparing AVP against a baseline (without an HMD) and an extensively used MR glass in the medical domain [14-22,24,25,28], MHL2, for performing 3 different suturing techniques. Our evaluation included both quantitative data, including the system usability score, task completion time (TCT), suturing performance, cognitive load, VR simulation sickness, and presence score, as well as qualitative data gathered through interviews.

Methods

Study Design and Protocol

This study adopts mixed methods research methodology, combining both quantitative and qualitative data collection approaches. Three interventions were designed, corresponding to the 3 conditions of the study: baseline (no HMD), MHL2, and AVP. The baseline (no HMD) condition was included to serve as a reference point for evaluating the effects of the other interventions. MHL2 was used due to its extensive prior use in the medical domain. A within-subject design was used, meaning that each participant participated in all studied interventions. The order of participation in each intervention was counterbalanced to mitigate potential order bias.

The study began with an introduction phase involving obtaining informed consent and an introduction to the study tasks from all participants. Following consent, the entire session was recorded using 2 cameras, one egocentric and one exocentric, with front-facing views. Later, participants were randomly assigned to 1 of 6 possible orders for performing the 3 interventions (baseline, MHL2, and AVP). Prior to task execution, participants completed 2 questionnaires on demographics and affinity for technology interaction [29].

Following completion of the preparation and order assignment, the experiment task began. During this stage, participants were asked to perform the same study task, which involved working with 3 different suture types, for each intervention. Prior to performing the study task with AVP and MHL2, calibration procedures were conducted. For MHL2, only eye calibration was performed. For AVP, both eye and hand calibration were performed to address any potential issues arising from lens misalignment or visual discrepancies.

After completing the task for each intervention, photographs of the participants' performance using the suturing kit were captured for subsequent evaluation. Additionally, web-based questionnaires were administered to evaluate key factors related to the user experience. Cognitive workload was assessed using

NASA Task Load Index (NASA-TLX) [30], VR-induced sickness was measured with the virtual reality sickness questionnaire (VRSQ) [31], and the sense of presence in the digital environment was evaluated using the presence questionnaire (PQ) [32]. Furthermore, for a rapid and reliable assessment of new health care technologies [33], UMUX-Lite questionnaire was used. System usability score was then predicted using a regression equation based on the 2 UMUX-Lite items [34]. The order of questions was presented in a random order for each participant. After completing the task for the baseline condition (without an HMD), participants completed only the NASA-TLX [30]. Following the MHL2 and AVP interventions, participants filled out the NASA-TLX, VRSQ, PQ, and UMUX-Lite.

Finally, at the end of each session, a researcher conducted a short semistructured interview with the participants, asking them to reflect on their experience with each device. The interview questions encapsulated aspects including comfort, self-performance evaluation, pros and cons, and potential use cases.

After all data recording sessions were completed, the TCTs were extracted from the recorded videos. A researcher, who was blinded to the study's aims, measured the TCT for each suture performed by every participant. To ensure an objective evaluation of the time spent solely on suturing, TCT was defined as the duration from the moment the needle was grasped by the needle holder until the knot was cut with scissors. Additionally, 5 surgeons (7.2 [SD 1.7] years of surgical experience), who were also unaware of the study aims, evaluated suture performance based on anonymized photographs of performed sutures. All of the captured photographs of participants' performed sutures were presented in a random order to evaluator surgeons using a custom visualization tool. They rated the performance of each suture type separately on a scale of 0 to 100, considering factors such as the overall effectiveness of the suture, bite (length of the stitch across the wound), pitch (interval between stitches), and cosmetic appearance [35].

Study Task

To assess the feasibility of using AVP while performing precision-dependent medical applications and compare it with MHL2 and baseline (no HMD), we designed a controlled user study task that incorporates performing different suture techniques. Since the main goal of this study was to evaluate the usability of the device and its underlying technology without any bias from specific applications, no digital information was displayed in the HMDs used in this study (AVP and MHL2). The participants were instructed to perform the suturing task either using one of the HMDs or without any HMD in the baseline condition. This approach is particularly important for delicate tasks requiring fine motor control and precision in real-world applications such as surgery, where the device itself may affect task performance regardless of the usability of the XR application used. By isolating the device from application-related factors, this ensured an unbiased assessment of its usability, preventing findings from being influenced by app design or content. The study included 3 types of sutures, each requiring progressively more complex techniques, ranging from basic to advanced. The simple interrupted suture (SIS, Figure 1A) was selected as the simplest task, while the vertical mattress suture (VMS, Figure 1B) and continuous subcuticular suture (CSS, Figure 1C) were chosen for their complexity. These techniques rely on correct depth perception, as they involve inserting a suturing needle into a specific layer of the skin [35], making them ideal for evaluating users' ability to perceive depth in a simulated environment.

A suture training kit [36] was used as the base for performing the sutures. All participants were asked to complete 3 SIS, 3 VMS, and close a 5 cm long wound on the suturing kit using a CSS for each intervention. All sutures were performed using 3‐0 polypropylene [37] and the same clinical surgical instruments, including needle holder, tweezers, and scissors (Figure 2). To minimize potential bias in TCTs due to the length of the suture material, each suture type was performed with a new suture material.

Figure 1. The illustration of 3 suture types included in the study task: (A) simple interrupted suture, (B) vertical mattress suture, and (C) continuous subcuticular suture.

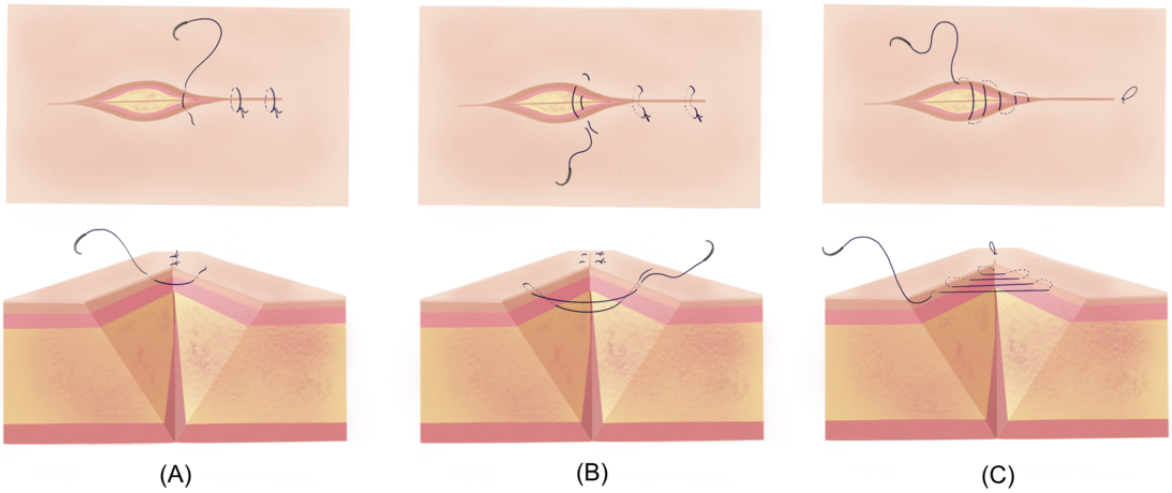
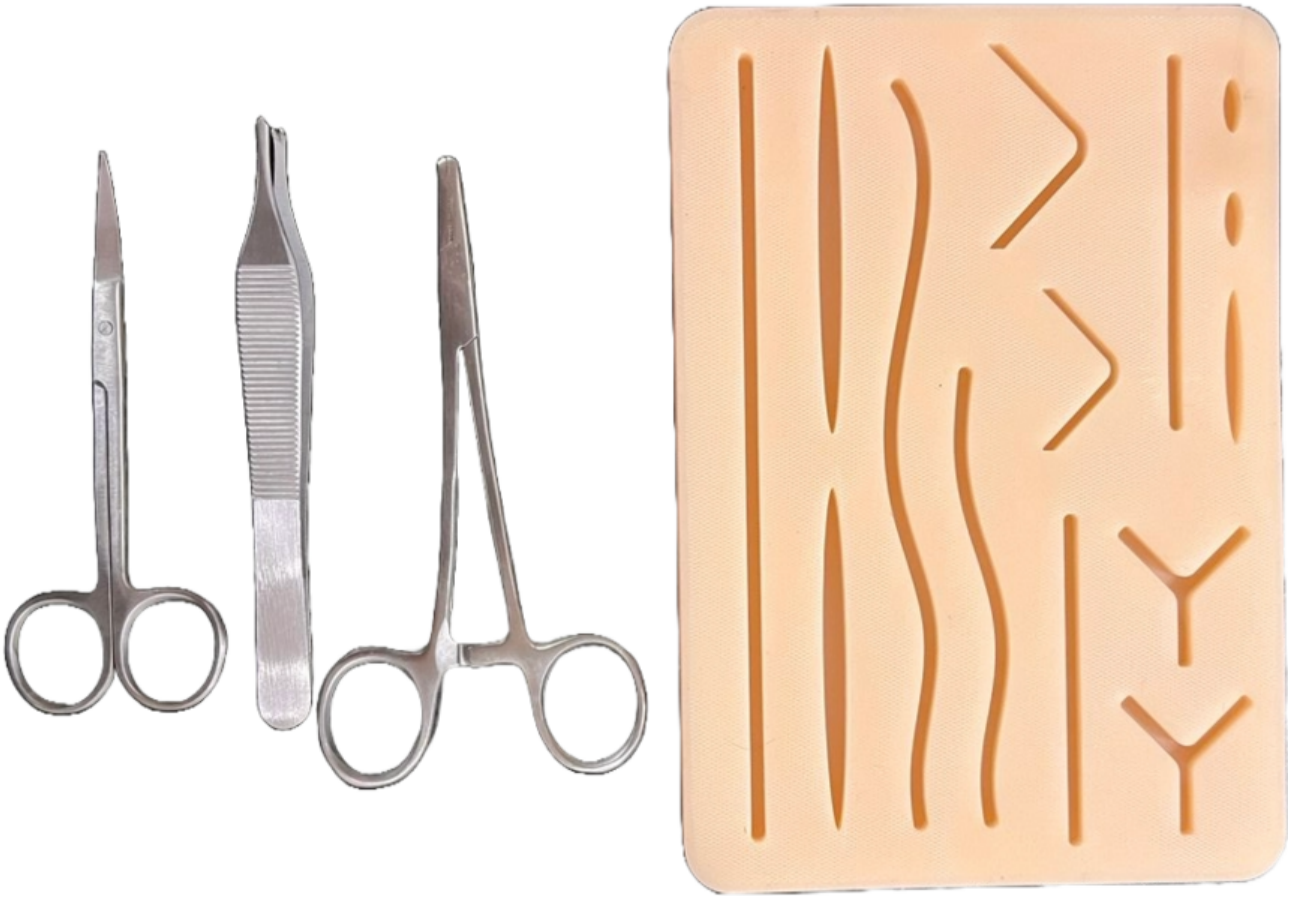


Figure 2. The suture training kit used during study tasks containing a silicone suture pad and instruments.



Participant Recruitment

Recruitment was conducted through word-of-mouth and advertisements via mail. The experiment took place at Klinikum Saarbrücken, Germany. Participation was entirely voluntary, and no compensation was provided.

The inclusion criteria for participants required that they be health care professionals with prior experience in performing wound suturing on patients. Additionally, the study limited participation to individuals without vision disorders or those with minor refractive errors, who could complete the study tasks without eyeglasses. To minimize bias, participants with minor refractive errors who typically used eyeglasses were instructed to perform all tasks without glasses across all interventions.

Statistical Analysis

Sample size calculation was performed using the power analysis tool G*Power [38]. Since there is no previous work comparing participants' performance using the AVP against MHL2 and baseline, we hypothesized a large between-group effect size (Cohen *f*) of 0.40 based on personal experience. This assumption was used to calculate the required sample size. With a power of $1–\beta=.95$ and $\alpha=.05$, the required sample size was calculated to be 18 participants. To account for 10% potential dropouts, we included a total of 20 participants in the study.

Statistical analysis was conducted using the R project for statistical computing [39]. Continuous data were expressed as means (SD), while categorical data were reported as frequencies and proportions. The effects of the interventions (baseline, MHL2, and AVP) were analyzed within-subject using repeated-measures ANOVA. When the assumption of sphericity was violated, the Greenhouse-Geisser ($\varepsilon<.75$) correction was applied. For the repeated measures ANOVA, we reported the F-statistic, degrees of freedom, *P* value, and generalized eta squared ($\eta G2$) as a measure of effect size. For post hoc pairwise comparisons, we conducted paired-samples *t* tests with Bonferroni correction for multiple within-group comparisons and reported the adjusted *P* values along with Cohen *d* to indicate effect size. A 2-sided *P* value of <.05 was considered statistically significant for all analyses. Furthermore, the inter-rater agreement of the performance scores by 5 evaluator surgeons was confirmed using the rwG(J) agreement index [40].

Qualitative Analysis

All interviews conducted in this project were transcribed verbatim. We adopted a pragmatic approach to qualitative analysis, as recommended by Blandford et al [41]. Initially, 2 researchers independently analyzed the same 25% of the data. Based on iterative discussions, a preliminary coding framework was developed. The remaining 75% of the interview data were then evenly distributed between the 2 researchers for coding using the established coding framework. To further ensure the consistency between final codings in case of new code emergence or coding disagreements, in a final discussion, the coding framework was further refined, leading to the development of the main themes.

Ethical Considerations

Before conducting the study, ethics approval was obtained from the institutional ethical review board of the German Research Center for Artificial Intelligence (DFKI, IRB approval number: VST – 48/25). All participants received comprehensive information about the objectives and data handling involved in this study. Data collection only proceeded after obtaining their voluntary informed consent. All participants were assured that their contributions would remain anonymous and were offered the opportunity to withdraw from the study at any stage prior to publication. All data were stored securely at the DFKI local server. Each participant signed a written consent form. No financial compensation was offered or provided.

Results

Study Population and Demographics

In total, 20 health care professionals participated in this study. The demographic characteristics of the participants are detailed in Table 1.

Table . Demographics of study participants.

Characteristics	Value
Gender, n (%)	
Man	14 (70)
Woman	6 (30)
Age (years), mean (SD)	33.65 (7.60)
Occupation, n (%)	
Surgeon	15 (75)
Physician assistant	1 (5)
Medical intern	4 (20)
Participants with minor refractive errors, n (%)	6 (30)
Clinical experience (years), mean (SD)	7.8 (6.45)
Prior use of OST[a] HMDs[b] (1 - 5 Likert scale), mean (SD)	1.8 (0.93)
Prior use of VST[c] HMDs (1 - 5 Likert scale), mean (SD)	1.35 (0.63)
Affinity for technology interaction (1 - 6 Likert scale), mean (SD)	3.83 (0.79)

[a]OST: optical see-through.

[b]HMD: head-mounted display.

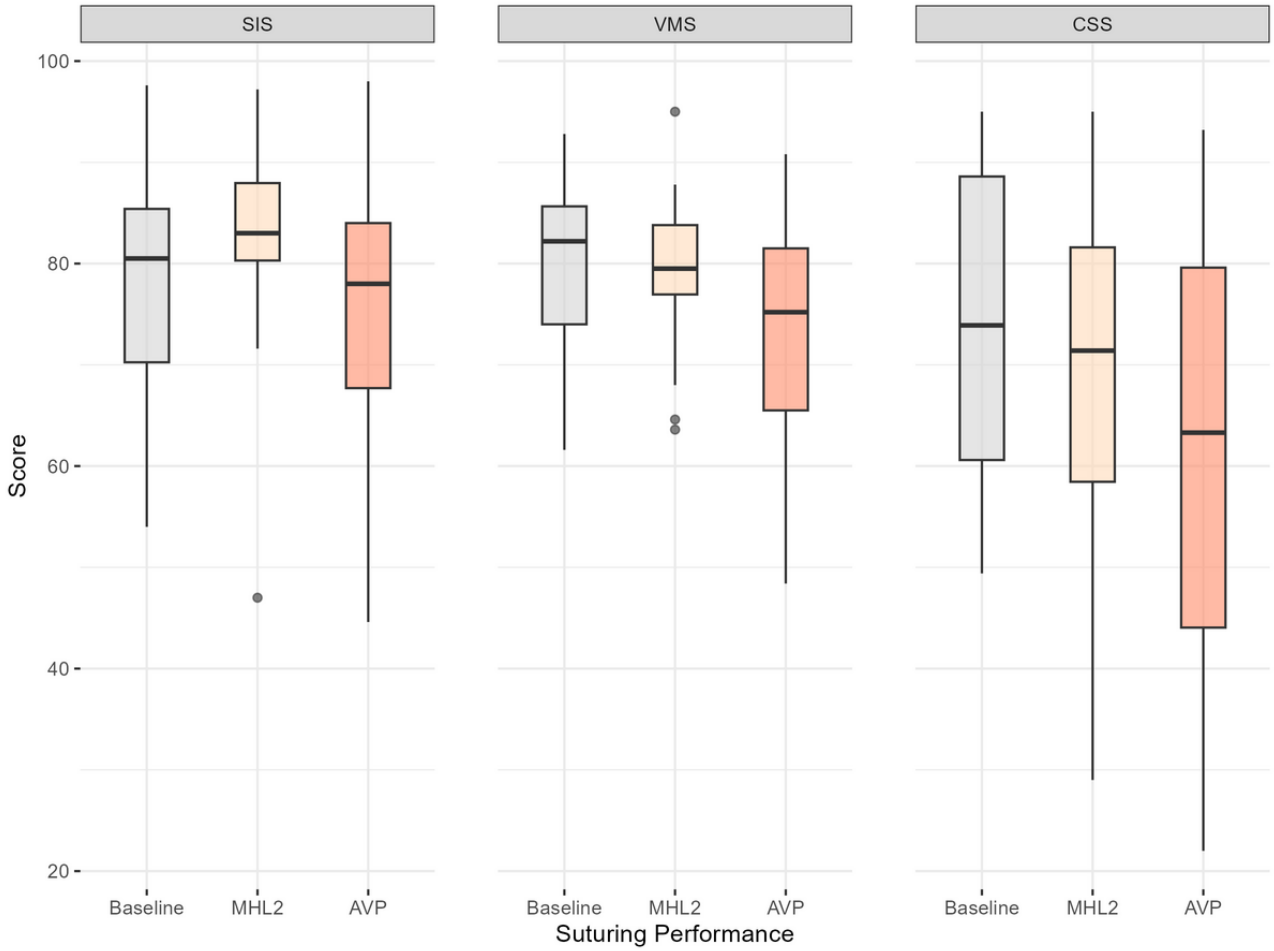
[c]VST: video-see-through.

User Performance

Suturing Performance

The 5 surgeons' evaluation scores of participants' suture performances showed high agreement, with an rwG(J) value greater than 0.99 across the suture types. The analysis on participants' average suture performance revealed no significant differences across interventions (baseline=73.8 [SD 13.5], AVP=74.5 [SD 10.0], MHL2=75.3 [SD 9.7], $F_{2, 38}$=0.28, P=.76, ηG2=0.003). Despite the worsened performance with AVP for all suture types, no significant differences were observed between interventions for any suture type (Figure 3).

Figure 3. The participants' suturing performance rated by surgeons across different suture types. AVP: Apple Vision Pro; CSS: continuous subcuticular suture; MHL: Microsoft HoloLens; SIS: simple interrupted suture; VMS: vertical mattress suture.



Task Completion Time

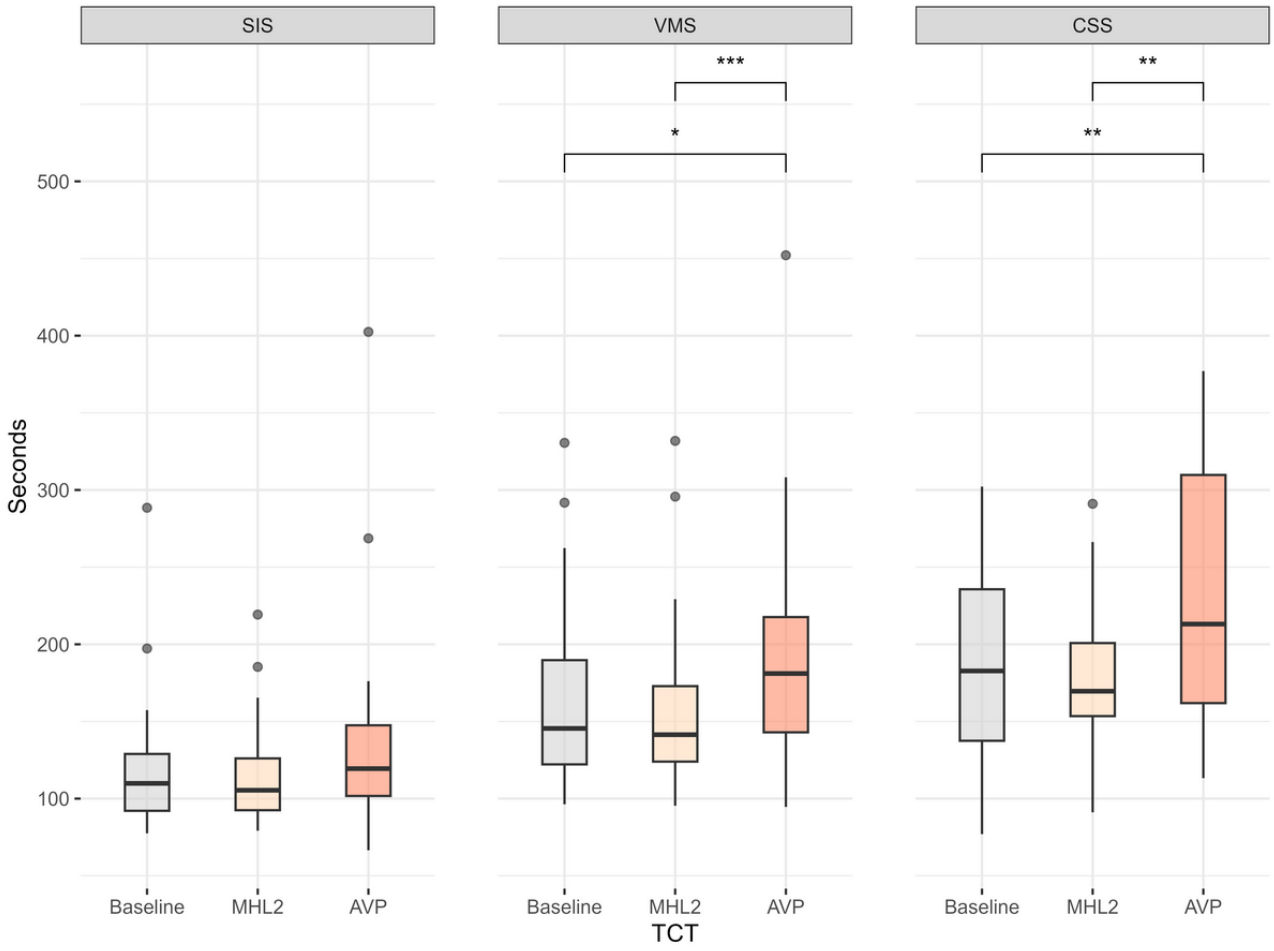
The analysis revealed a significant difference in the TCT required to complete all tasks across the interventions (baseline=472.0 [SD 143.0] s, AVP=570.0 [SD 192.0] s, MHL2=456.0 [SD 120.0] s; $F_{2, 38}$=17.6, *P*<.001, ηG2=0.101). Pairwise test results showed participants took significantly longer to complete the entire task using AVP to MHL2 (*P*<.001, Cohen *d*=1.02, large effect) and baseline (*P*<.001, Cohen *d*=1.03, large effect). The comparison between baseline and MHL2 showed a small effect size (*P*=.30, Cohen *d*=0.2, small effect), suggesting minimal difference.

Analysis performed on TCT for each individual suture type (Figure 4) revealed that there were no significant differences in the time required to perform SIS across the interventions (baseline=122.2 [SD 48.9] s, AVP=139.8 [SD 76] s, MHL2=117.5 [SD 37.5] s; $F_{1.4, 27.0}$=3.67, *P*=.05, ηG2=0.03). However, a significant difference was observed for more complex tasks, VMS (baseline=166.8 [SD 64.2] s, AVP =195.9 [SD 79.6] s, MHL2=161.2 [SD 63.4] s; $F_{1.5, 28.8}$=9.66, *P*=.001, ηG2=0.048) and CSS (baseline=182.9 [SD 59.1] s, AVP =234.7 [SD 83.6] s, MHL2=177.0 [SD 52.8] s; $F_{1.4, 27.3}$=11.7, *P*<.001, ηG2=0.138).

Pairwise test results showed participants required significantly more time to complete VMS using AVP compared with MHL2 (*P*<.001, Cohen *d*=1.0, large effect) and baseline (*P*=.04, Cohen *d*=0.6, moderate effect). The comparison between baseline and MHL2 showed no significant difference (*P*=.99, Cohen *d*=0.2, negligible effect).

Similarly, for CSS, a significantly longer time was needed when using AVP compared with MHL2 (*P*=.007, Cohen *d*=0.8, moderate effect) and baseline (*P*=.002, Cohen *d*=0.9, large effect). No significant difference was observed between baseline and MHL2 (*P*=.99, Cohen *d*=0.1, negligible effect).

Figure 4. Participants' TCTs across different suture types. Statistically significant differences are denoted as follows: *P<.05, **P<.01, and ***P<.001. AVP: Apple Vision Pro; CSS: continuous subcuticular suture; MHL: Microsoft HoloLens; SIS: simple interrupted suture; TCT: task completion time; VMS: vertical mattress suture.



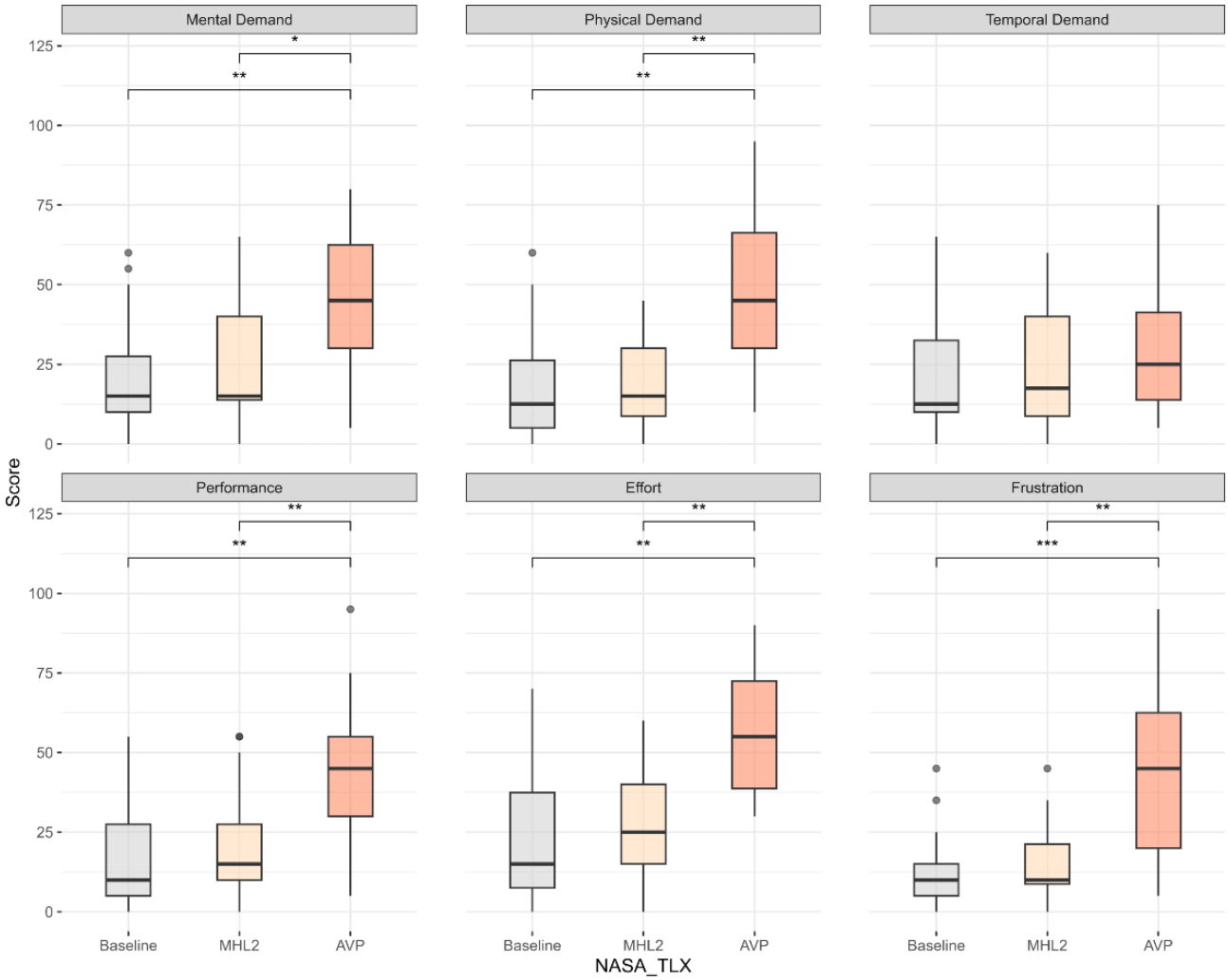
User Experience

Cognitive Workload

The analysis revealed a significant difference in the total raw NASA-TLX score across the interventions (baseline=19.1 [SD 15.1], AVP=43.9 [SD 15.9], MHL2=21.5 [SD 13.8]; $F_{1.2, 21.9}$=28.6 P<.001, ηG2=0.37). The pairwise test results showed the total raw NASA-TLX score among participants was significantly higher for AVP compared with MHL2 (P<.001, Cohen d=1.2, large effect) and baseline (P<.001, Cohen d=1.3, large effect). No significant difference was observed between baseline and MHL2 (P=.30, Cohen d=−0.401, negligible effect).

Analysis on each scale of NASA-TLX showed significant differences on mental demand (baseline=20.2 [SD 18.0], AVP=46.0 [SD 22.9], MHL2=23.75 [SD 18.6]; $F_{1.2, 22.8}$=21.1, P<.001, ηG2=0.256), physical demand (baseline=17.2 [SD 17.3], AVP =46.0 [SD 25.0], MHL2=19.2 [SD 14.3]; $F_{1.1, 21.3}$=22.4, P<.001, ηG2=0.324), performance (baseline =18.2 [SD 16.5], AVP =45.5 [SD 21.2], MHL2=20.7 [SD 16.6]; $F_{1.2, 22}$=3.6, P<.001, ηG2=0.3), effort (baseline=24.5 [SD 23.4], AVP =56.0 [SD 20.0], MHL2=28.0 [SD 18.0]; $F_{1.3, 24.5}$=23.4, P<.001, ηG2 =0.3), and frustration (baseline=12.25 [SD 11.7], AVP=43.5 [SD 23.4], MHL2=15 [SD 12.1]; $F_{1.1, 21}$=26.1, P<.001, ηG2=0.4) factors. Pairwise analysis showed significantly higher scores for AVP compared with MHL2 for mental demand (P<.01, Cohen d=1.0, large effect), physical demand (P<.001, Cohen d=1.2, large effect), performance (P<.001, Cohen d=1.0, large effect), effort (P<.001, Cohen d=1.1, large effect), and frustration (P<.001, Cohen d=1.1, large effect). Similarly, significant differences were observed for AVP compared with baseline on mental demand (P<.001, Cohen d=1.1, large effect), physical demand (P<.001, Cohen d=1.0, large effect), performance (P<.001, Cohen d=1.1, large effect), effort (P<.001, Cohen d=1.1, large effect), and frustration (P<.001, Cohen d=1.2, large effect). No significant difference was observed between MHL2 and baseline on any of the NASA-TLX factors (Figure 5).

Figure 5. NASA-TLX results for each factor (mental demand, physical demand, temporal demand, performance, effort, and frustration) across the 3 interventions. Statistically significant differences are denoted as follows: *P<.01, **P<.001, and ***P<.0001. AVP: Apple Vision Pro; MHL: Microsoft HoloLens; NASA-TLX: NASA Task Load Index.



Presence

The analysis of PQ demonstrated a significantly higher presence score (AVP=93.7 [SD 12.7], MHL2=115.0 [SD 11.4]; $F_{1, 19}$=27.9, P<.001, ηG2=0.4) for MHL2 compared with AVP. This significant difference was observed in all factors of presence including the realism (AVP=34.2 [SD 4.9], MHL2=41.0 [SD 5.3]; $F_{1, 19}$=17.2, P<.001, ηG2=0.3), possibility to act (AVP=20.6 [SD 3.5], MHL2=25.5 [SD 2.8]; $F_{1, 19}$=19.8, P<.001, ηG2=0.4), quality of interface (AVP=14.8 [SD 3.5], MHL2=18 [SD 3.0]; $F_{1, 19}$=19.7, P<.001, ηG2=0.2), possibility to examine (AVP=13.9 [SD 3.6], MHL2=18.3 [SD 2.3]; $F_{1, 19}$=20.1, P<.001, ηG2=0.4), and self-evaluation of performance (AVP=10.2 [SD 2.8], MHL2=12.2 [SD 2.0]; $F_{1, 19}$=5.5, P=.03, ηG2=0.1).

VR Sickness

The overall VRSQ score (AVP=66.9 [SD 19.8], MHL2=41.1 [SD 9.32]; $F_{1, 19}$=46.7, P<.001, ηG2=0.4) was also significantly higher for AVP compared with MHL2. The participants rated significantly higher scores on both oculomotor (AVP=75.8 [SD 22.9], MHL2=42.9 [SD 12.8]; $F_{1, 19}$=69.8 P<.001, ηG2=0.4) and disorientation (AVP=58 [SD 19.0], MHL2=39.3 [SD 9.15]; $F_{1, 19}$=18.9, P<.001, ηG20.3) factors for AVP compared with MHL2.

System Usability

System usability score for MHL2 was significantly higher compared with AVP (AVP=50.3 [SD 14.4], MHL2=72.7 [SD 8.54]; $F_{1, 19}$=34.4, P<.001, ηG2=0.5).

Interviews

Overview

After thematic analysis of the interviews, we developed 5 main themes: comfort and physical strain, visual challenges and depth perception, self-evaluation of performance, user confidence and preference, and application domain.

Comfort and Physical Strain

Thematic analysis of observations made during the study revealed several comfort-related issues associated with AVP. Participants reported discomfort due to the unbalanced weight distribution, with the majority of the weight concentrated on the nasal and maxillary area, leading to strain in the back neck

muscles and headaches. Two participants commented on this with the following statements:

> *When you look straight it is more convenient but when you bend your head to look at the stitch pad which I think would be the normal case when you operate around the table, then it is too uncomfortable, because the whole weight is in front and there is a constant contraction on your neck.* [P9, surgeon]

> *I usually get VR sickness whenever I use VR headsets. I tested AVP before and previously didn't have any issues when watching videos and so on, but here I got a very bad headache because I think I tried too much to focus and finish my task.* [P1, surgeon]

Additionally, they described a sensation of instability when focusing on a task for an extended period, expressing that it felt as though their head might fall forward.

> *It [AVP] is very heavy and after a while you feel like your head would fall down if you don't consistently fight it. And I can assure you it was a relief to take it out.* [P13, surgeon]

In contrast, no incidents of discomfort were observed with MHL2 for the period the participants used MHL2 for the experiment. One of the participants also indicated that the design of MHL2 is more comfortable and suited for use in the operating room compared with AVP.

> *We didn't use it here, but I guess with HoloLens you have also this option where you could push the visor up or down based on what you want to see but with AVP you don't even have that option. You just have to take it off completely. And it is just not practical to use it during operation if you have to take it off every time.* [P1, surgeon]

Visual Challenges and Depth Perception

Participants also experienced visual challenges with AVP, including blurred vision and difficulties with depth perception. The struggle with depth perception further impacted their ability to accurately judge distances, which is critical for suturing.

> *My vision felt a bit blurry; I could not see the details I had difficulties to see the needle and also to do the knots.* [P12, surgeon]

> *For the first two sutures AVP was also ok but for the last suture type you have to really see where you put your needle in and that was simply impossible to make sure you are in the correct layer.* [P11, physician assistant]

> *I don't think that my hand and eye coordination was disturbed, but it was very difficult to estimate the depth, there was like less contrast compared to reality.* [P4, surgeon]

Participants also reported that the sharpness of the view varied depending on the distance of the objects from them. Some noted that maintaining a greater distance provided a sharper view; however, this was not ideal for delicate tasks, as they naturally tended to lean in for a closer observation and better precision.

> *I could see things clearly farther than one meter to me. I could even read small letters, but when I looked at the stitching pad which was closer to me then it became blurry. And got even blurrier when I was leaning closer to it to do the stitches. Which usually you should see better when you get closer, but it just made it worse.* [P13, surgeon]

Self-Evaluation of Performance

Some participants believed that AVP negatively impacted their performance compared with MHL2 and baseline.

> *I used AVP before, but I only used it to watch videos, and initially I thought AVP would be better compared to HoloLens, but it was a complete catastrophe. I almost saw nothing. Yes, I did sutures from experience, but it was a complete guess work especially for the last suture type.* [P1, surgeon]

> *Naturally you realize a difference between no glass and having glass for both devices. But the difference was simply too much for AVP that I think it really impacted my performance.* [P11, physician assistant]

> *The HoloLens was not a big influence in my performance compared to performing without one it felt like having a light shaded sunglass on. If you look through the screen the vision is a bit darker but doesn't make your performance worse. AVP is very immersive but for precise work like stitches it's not fast enough and the resolution is not optimal.* [P4, surgeon]

User Confidence and Preference

Participants expressed a preference for OST over VST. They reported that VST created a sense of disconnection from the real world, which was also evident in the lower presence score compared with OST. They reported that in real-life scenarios involving patients, it would impact their confidence. In contrast, OST allowed them to maintain situational awareness and benefit from a wider peripheral vision, enhancing their overall experience and performance. Participants reflected on this, saying:

> *With HoloLens I felt more secure, because I think my peripheral vision was not affected that much but in AVP even though you still see but you have more restricted peripherals.* [P2, surgeon]

> *I think optimally the best is no glass but if I should choose, I think when you talk to a patient or your colleagues around the surgery table it feels just more natural to have eye contact even if it is through a glass like this [showing MHL2]. It is better than having a big headset on your face where no one can see your eyes in there. It is just more assuring with a see-through glass than a completely closed one [AVP].* [P18, surgeon]

Application Domain

The participants regarded MHL2 as a usable device for various medical applications, including applications for intraoperative use. However, they believed that AVP would be more suitable

for domains such as training or surgical planning, where the device would not be used during actual patient operations.

> *I think both devices could be used for medical domain, but I won't feel comfortable operating with the first one [AVP]. I think it is risky if you operate on veins or arteries. I don't want to take any risk when operating on a patient.* [P2, surgeon]

> *I see HoloLens as a usable device during operation, it won't stress you, but AVP would be a better fit for training or teaching or perhaps surgery planning.* [P8, surgeon]

A participant also suggested that AVP could be used for surgical applications, such as laparoscopic surgery, where the surgical field is already viewed through high-resolution video.

> *I think AVP would be useful for laparoscopic surgery where your view to the operation scene is already through a video and with AVP you can have this high-quality video stream.* [P1, surgeon]

Discussion

Overview

With advancements in computational power, camera technology, and display systems, a noticeable trend is emerging in commercially produced HMDs. Manufacturers are increasingly shifting from dedicated AR and VR HMDs toward XR HMDs capable of supporting both functionalities. This transition is also evident in the evolution of recent HMDs developed by well-known brands, such as Apple [3] and Meta [42], which use VST displays, in contrast to earlier designs like Microsoft's HoloLens [13], which relied on OST technology. Although this transition is expected to bring advantages beyond simply combining AR and VR, such as a wider field of view, higher camera quality, brightness control, and ultimately more precise spatial registration of digital objects, its feasibility in domains requiring high precision, such as intraoperative use, remains untested. Despite the foreseen potential benefits that recent XR glasses could bring to the medical domain [4,5,12], their feasibility in high-precision medical applications remained an open question. While most related studies focus on evaluating specific immersive applications within this domain [7,28,43], the choice of used devices is often driven by market trends rather than a critical assessment of their suitability for the intended use. Although some studies have compared the technical capabilities of different HMDs, including VST and OST design [44-46], they often overlook user experience and performance outcomes.

The findings of this study underscore the importance of device selection, particularly in time-sensitive and precision-dependent medical contexts. As progressively more immersive HMDs are being produced, our results demonstrate that the appropriateness of the chosen device itself plays a pivotal role in the user's real-world performance—even before any application is introduced. Neglecting to assess the suitability of the device as an initial step may contribute to the negative user experience and delayed integration of immersive technologies in clinical settings, as the hardware itself may be ill-suited for the domain—even when the application might offer substantial potential benefits.

Principal Results

In this study, we evaluated the feasibility of AVP (the most recent and promising XR HMD for the medical domain [10,12]) with MHL2, the commonly used MR HMD in the medical domain for precision tasks [14-22,24,25,28]. Twenty health care professionals participated in the study, performing suture tasks under 3 conditions: AVP, MHL2, and a baseline condition without an HMD. We evaluated user performance and experience across these conditions. Post hoc analysis of the primary outcome measures revealed a large observed effect size between AVP, MHL2, and the baseline, suggesting that the sample size (n=20) was adequate and confirming the validity of our initial power analysis. Our findings highlight the distinct strengths of each HMD, suggesting their suitability for different medical use cases.

A key observation from our study was the significant increase in cognitive load when using AVP compared with MHL2 and the baseline condition. This higher cognitive burden was accompanied by increased TCTs, indicating that AVP demands a greater cognitive workload from users. Interestingly, despite these challenges, objective performance evaluations did not show significant differences across the interventions. However, participants' self-reported performance, as captured in the PQ, indicated a significant decline compared with MHL2. This divergence between expert-assessed performance scores and self-evaluations may reflect a reduction in participants' confidence in their task performance, a sentiment that was also echoed in the interview responses.

These findings suggest that AVP remains a feasible option for non-time-sensitive medical domains where cognitive workload and TCT are not critical factors, such as a surgical planning tool demonstrated by Olexa et al [7], where there is no trade-off between the benefits of the used application and the need for real-world precision and speed.

Moreover, AVP potentially stands as a feasible device for medical education and training, incorporating virtual reality simulations where interaction with digital elements is prioritized over real-world precision. However, the usability of AVP while training precision tasks in the real world might be limited. A participant's sentiment on performing sutures from experience instead of relying on their visual perceptions indicates that training real-world precision tasks with AVP could potentially cause negative impacts on the learning process. A device that inadvertently increases cognitive demand or induces disorientation, as observed with AVP (Figure 5), might lead to suboptimal skill acquisition, potentially compromising the training outcomes. In educational contexts, the choice of XR technology can significantly influence learning behavior. Although positive outcomes have been demonstrated in nonmedical areas such as design education [47] over nonimmersive devices, the choice of immersive devices should still be carefully considered in educational settings, particularly those that emphasize psychomotor skills. An educational XR system that imposes excessive cognitive workload or fails to foster a strong sense of presence may lead to the development

of maladaptive motor patterns or “incorrect” muscle memory, ultimately impacting long-term real-world performance.

Furthermore, AVP could serve as a practical solution for applications where direct real-world perception is either not required or is transmitted to the device instead of being captured by it, such as in telemedicine. Similarly, in fields where user presence is essential but critical information is traditionally delivered through digital interfaces, like laparoscopic surgery, where surgeons view the operative field on a 2D monitor, the potential of AVP could be further explored and leveraged. However, for use cases that involve fine motor tasks, such as those suggested by previous works including medical training scenarios [10,48] or in more critical applications such as intraoperative support tools [48], the increased workload and disorientation associated with AVP might hinder user real-world performance. Conversely, MHL2 demonstrated its suitability for applications requiring high precision and time efficiency. Its OST design allowed for a lower cognitive load and better user experience during the suturing task, with no significant differences when compared with the baseline. This underscores its potential for intraoperative use cases, such as surgical navigation and other real-time assistance tools, where maintaining a strong connection to the real world is critical.

Unlike the findings of the study by Olexa et al [7], where participants reported minimal eye strain or fatigue, our questionnaire responses on VR sickness indicated heightened oculomotor strain and disorientation with AVP. This discrepancy could be attributed to differences in user attention directed toward digital versus real-world objects. In the study by Olexa et al [7], users primarily focused on digital objects, whereas in our study, the main focus was on real-world perception. This phenomenon was also corroborated by one of our participants, who had prior experience using AVP. In our study, participants frequently reported that AVP caused physical discomfort—including eye strain, headaches, and neck fatigue—as well as visual challenges such as blurriness, difficulty focusing, and disorientation during head movement. These sensory and ergonomic limitations often led to reduced task confidence and greater reliance on prior experience or instinct rather than real-time visual feedback. In contrast, MHL2 was consistently described as lightweight, comfortable, and minimally intrusive, with clearer visual output and fewer disruptions to the user’s natural workflow. Such qualitative insights underscore the critical role of comfort and visual clarity in sustaining task engagement and motor coordination over time. These human-centered considerations are especially relevant in domains where extended use or precision is essential. AVP, in its current form with display quality and ergonomic constraints, may be more suitable for fully immersive VR applications rather than AR/MR-integrated medical use, particularly for shorter durations. In contrast, MHL2 demonstrated lower VR sickness scores, attributed to its OST design, along with a more balanced weight distribution and higher presence scores. These features position MHL2 as better optimized for applications requiring extended use periods and seamless real-world connectivity.

The findings of this study highlight several factors that could inform the future design of XR and MR devices for the medical domain. While generalizability remains a goal for widespread adoption across diverse medical applications, custom designs may be more appropriate for time-critical use cases, such as surgical navigation systems. The optimal approach may vary depending on the method used to visualize the surgical scene, whether open surgery, laparoscopic, or robotic-assisted procedures. For open surgery, OST displays could provide distinct advantages by preserving a clear view of the real world and facilitating seamless communication with the surgical team. Conversely, in scenarios involving indirect surgical views, such as laparoscopic procedures, VST HMDs might offer greater benefits. Furthermore, the future integration of VST HMDs in precision-demanding applications requires high-quality camera feed from a real-world environment; challenges such as camera focus issues, which result in blurring of the real-world video stream, can significantly hinder the usability of these devices in tasks requiring precision. Finally, ensuring comfort during extended use is critical for intraoperative tools. Features such as balanced weight distribution, antimicrobial coatings, and easy-to-clean surfaces would further facilitate smoother integration into clinical workflows.

Limitations

While this study provides valuable insights, it also has several limitations. First, we focused solely on evaluating the feasibility of AVP for medical precision tasks and compared the outcomes with MHL2 as a representative example of existing MR devices, given its extensive prior use in the medical domain. To further validate the generalizability of our findings regarding comparisons between various VST and OST displays, additional research involving other available HMDs is necessary. Second, our study included only participants with no or minor refractive errors who were able to complete the tasks without eyeglasses. To minimize bias, the 6 participants with minor refractive errors were asked to perform all tasks across all interventions without wearing eyeglasses. Although it is possible to wear eyeglasses with MHL2, this approach was not feasible for AVP due to its design. While there is an option to integrate correction lenses into AVP, customizing lenses for each participant was impractical and not feasible. Finally, while AVP is expected to enhance the display of digital objects, no digital elements were incorporated into the study tasks, as this was beyond the scope of our research. Our primary objective was to assess the safety and feasibility of AVP as a VST-HMD for performing medical precision tasks.

Conclusions

In conclusion, while AVP shows promise for non-time-sensitive applications that do not have an emphasis on real-world perception, MHL2 remains the preferred choice for time-critical and precision-demanding tasks. Further research and device refinements will be necessary to fully integrate XR HMDs into diverse medical applications, ensuring both user comfort and operational efficiency.

Acknowledgments

We acknowledge the financial support provided by the European Union's Horizon 2020 program under award number 952026, HumanE-AI-Net, and BMBF funding in the context of the VidgenSense project, which made this work feasible.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors have made substantial contributions to the article to meet the criteria for authorship. HJ and OG were responsible for the conceptualization of the study. The methodology was developed by HJ, PL, GAS, JK, and OG. Data recording and curation were conducted by HJ, VFR, and OG. Formal analysis was carried out by HJ, VFR, JK, and OG. PL and JK were responsible for funding acquisition. Supervision was provided by PL, GAS, JK, and OG. The original draft of the manuscript was written by HJ, VFR, and OG. All authors—HJ, VFR, PL, GAS, JK, and OG—contributed to the review and editing of the manuscript.

Conflicts of Interest

None declared.

References

1. Milgram P, Takemura H, Utsumi A, Kishino F. Augmented reality: a class of displays on the reality-virtuality continuum. Presented at: Telemanipulator and Telepresence Technologies; Oct 31-4, 1994; Boston, Massachusetts, USA.
2. Dargan S, Bansal S, Kumar M, Mittal A, Kumar K. Augmented reality: a comprehensive review. Arch Computat Methods Eng 2023 Mar;30(2):1057-1080. [doi: 10.1007/s11831-022-09831-7]
3. Apple Vision Pro. Apple Inc. URL: https://www.apple.com/de/shop/buy-vision/apple-vision-pro [accessed 2025-07-31]
4. Egger J, Gsaxner C, Chen X, Bian J, Kleesiek J, Puladi B. Apple Vision Pro for healthcare: "the ultimate display"? -- entering the wonderland of precision medicine. JMIR Preprints. Preprint posted online on Sep 15, 2023. [doi: 10.2196/preprints.52785]
5. Egger J, Gsaxner C, Luijten G, et al. Is the Apple Vision Pro the ultimate display? a first perspective and survey on entering the wonderland of precision medicine. JMIR Serious Games 2024 Sep 18;12(1):e52785. [doi: 10.2196/52785] [Medline: 39292499]
6. Masalkhi M, Waisberg E, Ong J, et al. Apple Vision Pro for ophthalmology and medicine. Ann Biomed Eng 2023 Dec;51(12):2643-2646. [doi: 10.1007/s10439-023-03283-1] [Medline: 37332003]
7. Olexa J, Trang A, Cohen J, et al. The Apple Vision Pro as a neurosurgical planning tool: a case report. Cureus 2024 Feb;16(2):e54205. [doi: 10.7759/cureus.54205] [Medline: 38496193]
8. Orione M, Rubegni G, Tartaro R, et al. Utilization of Apple Vision Pro in ophthalmic surgery: a pilot study. Eur J Ophthalmol 2025 Mar;35(2):715-721. [doi: 10.1177/11206721241273574] [Medline: 39140319]
9. Santos E, Castillo V. Apple vision pro: comments in healthcare. arXiv. Preprint posted online on Jan 13, 2024. [doi: 10.48550/arXiv.2401.08685]
10. Shanbhag NM, Bin Sumaida A, Al Shamisi K, Balaraj K. Apple Vision Pro: a paradigm shift in medical technology. Cureus 2024 Sep;16(9):e69608. [doi: 10.7759/cureus.69608] [Medline: 39308843]
11. Tsai YC, Hsiao CC, Chung-Wei Lin C, Chern MC, Huang SW. Apple Vision Pro-guided laparoscopic radio frequency ablation for liver tumors: the pioneer experience. Surg Innov 2025 Jun;32(3):312-314. [doi: 10.1177/15533506251316001] [Medline: 39853152]
12. Waisberg E, Ong J, Masalkhi M, et al. Apple Vision Pro and why extended reality will revolutionize the future of medicine. Ir J Med Sci 2024 Feb;193(1):531-532. [doi: 10.1007/s11845-023-03437-z] [Medline: 37365445]
13. Microsoft HoloLens. Microsoft. URL: https://learn.microsoft.com/de-de/hololens/ [accessed 2025-07-31]
14. Orchard L, Van M, Abbas J, Malik R, Stevenson J, Tolley N. Mixed-reality technology for clinical communication: objective assessment of the HoloLens 2 as a clinical communication device in a simulated on-call scenario. J Laryngol Otol 2023 Oct;137(10):1165-1169. [doi: 10.1017/S0022215123000531]
15. Hilt AD, Hierck BP, Eijkenduijn J, et al. Development of a patient-oriented Hololens application to illustrate the function of medication after myocardial infarction. Eur Heart J Digit Health 2021 Sep;2(3):511-520. [doi: 10.1093/ehjdh/ztab053] [Medline: 36713611]
16. Shah SHH, Longva B, Hameed IA, Solberg M, Susanne TKA. Health data management for nursing practice: an intelligent, holographic mixed-reality system. Presented at: HCI International 2021-Posters: 23rd HCI International Conference, HCII 2021, Virtual Event; Jul 24-29, 2021.
17. Sugimoto M, Sueyoshi T. Development of Holoeyes holographic image-guided surgery and telemedicine system: clinical benefits of extended reality (virtual reality, augmented reality, mixed reality), the metaverse, and artificial intelligence in surgery with a systematic review. MRAJ 2023;11(7.1). [doi: 10.18103/mra.v11i7.1.4045]

18. Kumar N, Pandey S, Rahman E. A novel three-dimensional interactive virtual face to facilitate facial anatomy teaching using Microsoft HoloLens. Aesthetic Plast Surg 2021 Jun;45(3):1005-1011. [doi: 10.1007/s00266-020-02110-5] [Medline: 33469701]
19. Moro C, Phelps C, Redmond P, Stromberga Z. HoloLens and mobile augmented reality in medical and health science education: a randomised controlled trial. Br J Educ Technol 2021 Mar;52(2):680-694. [doi: 10.1111/bjet.13049]
20. Javaheri H, Gruenerbl A, Monger E, Gobbi M, Karolus J, Lukowicz P. RescuAR: a self-directed augmented reality system for cardiopulmonary resuscitation training. Presented at: 17th EAI International Conference on Pervasive Computing Technologies for Healthcare; Nov 27-29, 2023; Malmö, Sweden.
21. Gibby JT, Swenson SA, Cvetko S, Rao R, Javan R. Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. Int J Comput Assist Radiol Surg 2019 Mar;14(3):525-535. [doi: 10.1007/s11548-018-1814-7] [Medline: 29934792]
22. Javaheri H, Ghamarnejad O, Bade R, Lukowicz P, Karolus J, Stavrou GA. Beyond the visible: preliminary evaluation of the first wearable augmented reality assistance system for pancreatic surgery. Int J Comput Assist Radiol Surg 2025 Jan;20(1):117-129. [doi: 10.1007/s11548-024-03131-0] [Medline: 38849631]
23. Javaheri H, Ghamarnejad O, Lukowicz P, Stavrou GA. LLMs enable context-aware augmented reality in surgical navigation. Presented at: DIS '25: Proceedings of the 2025 ACM Designing Interactive Systems Conference; Jul 5-9, 2025; Madeira, Funchal. [doi: 10.1145/3715336.3735708]
24. Javaheri H, Ghamarnejad O, Lukowicz P, Stavrou GA, Karolus J. Design and clinical evaluation of ARAS: an augmented reality assistance system for open pancreatic surgery. Presented at: 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); Oct 21-25, 2024; Bellevue, WA, USA. [doi: 10.1109/ISMAR62088.2024.00052]
25. Javaheri H, Ghamarnejad O, Widyaningsih R, et al. Enhancing perioperative outcomes of pancreatic surgery with wearable augmented reality assistance system: a matched-pair analysis. Ann Surg Open 2024 Dec;5(4):e516. [doi: 10.1097/AS9.0000000000000516] [Medline: 39711676]
26. Liu X, Sun J, Zheng M, Cui X. Application of mixed reality using optical see-through head-mounted displays in transforaminal percutaneous endoscopic lumbar discectomy. Biomed Res Int 2021;2021:9717184. [doi: 10.1155/2021/9717184] [Medline: 33644234]
27. Sirilak S, Muneesawang P. A new procedure for advancing telemedicine using the HoloLens. IEEE Access 2018;6:60224-60233. [doi: 10.1109/ACCESS.2018.2875558]
28. Gsaxner C, Li J, Pepe A, et al. The HoloLens in medicine: a systematic review and taxonomy. Med Image Anal 2023 Apr;85:102757. [doi: 10.1016/j.media.2023.102757] [Medline: 36706637]
29. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. Int J Hum Comput Interact 2019 Apr 3;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
30. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Advances in Psychology: Elsevier; 1988:139-183.
31. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. Appl Ergon 2018 May;69:66-73. [doi: 10.1016/j.apergo.2017.12.016] [Medline: 29477332]
32. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence (Camb) 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
33. Borsci S, Buckle P, Walne S. Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology? Appl Ergon 2020 Apr;84:103007. [doi: 10.1016/j.apergo.2019.103007] [Medline: 31785449]
34. Lewis JR, Utesch BS, Maher DE. Investigating the correspondence between UMUX-LITE and SUS scores. design, user experience, and usability: design discourse. Presented at: 4th International Conference, DUXU 2015, Held as Part of HCI International 2015; Aug 2-7, 2015; Los Angeles, CA, USA. [doi: 10.1007/978-3-319-20886-2_20]
35. Latona JA, Tannouri S, Palazzo F, Pucci MJ. Fundamentals of sutures, needles, knot tying, and suturing technique. In: Fundamentals of General Surgery: Springer Cham; 2018:16-350. [doi: 10.1007/978-3-319-75656-1_3]
36. Mediarchitect Suture Kit. Mediarchitect. URL: https://mediarchitect.net/en-de/collections/suture-practice [accessed 2025-07-31]
37. PROLENE nahtmaterial. Johnson & Johnson MedTech. URL: https://www.jnjmedtech.com/de-DE/product/prolene-nahtmaterial [accessed 2025-07-31]
38. G*Power: statistical power analyses for Mac and Windows. Heinrich-Heine-Universität Düsseldorf. URL: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower [accessed 2025-07-31]
39. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2025-07-31]
40. James LR, Demaree RG, Wolf G. Estimating within-group interrater reliability with and without response bias. J Appl Psychol 1984;69(1):85-98. [doi: 10.1037/0021-9010.69.1.85]
41. Blandford A, Furniss D, Makri S. Qualitative HCI Research: Going behind the Scenes: Morgan & Claypool Publishers; 2016.
42. Meta Quest. Meta. URL: https://www.meta.com/de/quest [accessed 2025-07-31]

43. Li C. Apple Vision Pro as an augmentative tool in surgery: a double-edged scalpel. Ann Med Surg 2024;86(11):6407-6408. [doi: 10.1097/MS9.0000000000002645]
44. Soares I, B Sousa R, Petry M, Moreira AP. Accuracy and repeatability tests on HoloLens 2 and HTC Vive. Multimodal Technol Interact 2021;5(8):47. [doi: 10.3390/mti5080047]
45. Hu T, Yang F, Scargill T, Gorlatova M. Apple v.s. meta: a comparative study on spatial tracking in SOTA XR headsets. Presented at: ACM MobiCom '24: Proceedings of the 30th Annual International Conference on Mobile Computing and Networking; Nov 18-22, 2024; Washington DC, USA. [doi: 10.1145/3636534.3696215]
46. Cheng L, Schreiner M, Kunz A. Comparing tracking accuracy in standalone MR-HMDs: Apple Vision Pro, HoloLens 2, Meta Quest 3, and Pico 4 Pro. Presented at: VRST '24: Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology; Oct 9-11, 2024; Trier, Germany. [doi: 10.1145/3641825.3689518]
47. Xia L, Li X, Qin Y, Li D, Fan L. Enhancing design education through spatial computing: a comparative study of traditional and immersive technologies in chair design projects. Presented at: 2024 IEEE Frontiers in Education Conference (FIE); Oct 13-16, 2024; Washington, DC, USA. [doi: 10.1109/FIE61694.2024.10893086]
48. Waisberg E, Ong J, Masalkhi M, et al. Apple Vision Pro and the advancement of medical education with extended reality. Can Med Educ J 2024 Feb;15(1):89-90. [doi: 10.36834/cmej.77634] [Medline: 38528897]

Abbreviations

AR: augmented reality
AVP: Apple Vision Pro
CSS: continuous subcuticular suture
HMD: head-mounted display
MHL: Microsoft HoloLens
MR: mixed reality
NASA-TLX: NASA Task Load Index
OST: optical see-through
PQ: presence questionnaire
SIS: simple interrupted suture
TCT: task completion time
VMS: vertical mattress suture
VR: virtual reality
VRSQ: virtual reality sickness questionnaire
VST: video-see-through
XR: extended reality

Edited by L Riedemann; submitted 07.03.25; peer-reviewed by A Adeoye, P Boffi; revised version received 09.06.25; accepted 11.07.25; published 15.08.25.

Please cite as:
Javaheri H, Fortes Rey V, Lukowicz P, Stavrou GA, Karolus J, Ghamarnejad O
Assessing the Feasibility of Using Apple Vision Pro While Performing Medical Precision Tasks: Controlled User Study
JMIR XR Spatial Comput 2025;2:e73574
URL: https://xr.jmir.org/2025/1/e73574
doi:10.2196/73574

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org