Methodology for Clinical Trials of Virtual Reality in Healthcare: Recommendations from an International Working Group

Brandon Birckhead, MD; Carine Khalil, PhD; Xiaoyu Liu, MPH; Samuel Conovitz, MPH; Albert Rizzo, PhD; Itai Danovitch, MD, MBA; Kim Bullock, MD; and Brennan Spiegel, MD, MSHS

On behalf of the Virtual Reality Committee of Outcomes Research Experts (VR-CORE)

(1) Cedars-Sinai Health System, Division of Health Services Research, Department of Medicine; (2) Paris Descartes University, Department of Management Science; (3) UCLA Fielding School of Public Health, Department of Health Policy and Management; (4) University of Southern California Institute for Creative Technologies; (5) Cedars-Sinai Health System, Department of Psychiatry; (6) Stanford University School of Medicine, Department of Psychiatry

Running Title: VR Clinical Trial Methods
Abstract Word Count: 249
Manuscript Word Count: 3800
Grant Support: This work is unfunded.
Number of Figures and Tables: 3 tables

Corresponding Author Contact Information
Brennan Spiegel, MD MSHS
Director of Health Services Research, Cedars-Sinai Health System
Professor of Medicine and Public Health in Residence
Pacific Theatre Building
116 N. Robertson Blvd, 8th Floor
Los Angeles, Calif. 90048
310.324.6467
brennan.spiegel@cshs.org
Twitter: @BrennanSpiegel
ABSTRACT

Background: Therapeutic virtual reality (VR) has emerged as an efficacious treatment modality for a wide range of health conditions. However, in spite of encouraging outcomes from early stage research, a consensus is needed for how best to develop and evaluate VR treatments within a scientific framework.

Objective: We sought to develop a methodological framework with input from an international working group to guide the design, implementation, analysis, interpretation, and communication of trials that develop and test VR treatments.

Methods: A group of 21 international experts was recruited based upon contributions to the VR literature. The resulting Virtual Reality Committee of Outcomes Research Experts (VR-CORE) held iterative meetings to seek consensus regarding best practices for development and testing of VR treatments.

Results: The interactions were transcribed, and key themes were identified to support a scientific framework to support methodology best practices for clinical VR trials. Using the Food and Drug Administration (FDA) Phase I-III pharmacotherapy model as guidance, a framework emerged to support three phases of VR clinical study designs, herein named VR1, VR2, and VR3. VR1 studies focus on content development by working with patients and providers through principles of human centered design. VR2 trials conduct early testing with a focus on feasibility, acceptability, tolerability and initial clinical efficacy. VR3 trials are randomized, controlled studies to evaluate efficacy versus a control condition. Best practice recommendations for each trial are provided.

Conclusion: Patients, providers, payers and regulators may consider this best practice framework when assessing the validity of VR treatments.

Keywords: Virtual Reality, Clinical Trials, Consensus
Therapeutic virtual reality (VR) is an innovative treatment modality gaining considerable attention to manage a broad range of health conditions\[1–19\]. Users of VR wear a head-mounted display (HMD) with a close-proximity screen that creates a sensation of being transported into lifelike, three-dimensional worlds. VR has been used to assess and treat a wide variety of medical, surgical, psychiatric, and neurocognitive conditions including pain\[1,2,4,9,13,18\], addiction\[20–25\], anxiety disorders\[3,6,7,15,26–35\], schizophrenia\[10,11,36–40\], eating disorders\[1,41–48\], stroke rehabilitation\[5,16,48–51\], vestibular disorders\[52\], and movement disorders\[53\]. There have also been fMRI studies demonstrating the effect of VR on the brain while undergoing a painful stimuli\[54,55\]. VR is thought to work through a combination of distraction, extinction learning, cognitive-behavioral principles, mindful meditation, stress reduction, gate-control theory, and the spotlight theory of attention\[56,57\]. Importantly, VR has become increasingly portable, immersive and vivid, which has enabled the technology to be used in a broad range of inpatient and outpatient applications.

As the use of therapeutic VR expands, it is essential that guidelines are established to ensure scientific rigor in the development and evaluation of VR applications, similar to established standards for pharmacotherapies\[30,58\]. VR developers would benefit from systematic guidance regarding best practices for the design and conduct of VR clinical trials. To fill this unmet need, we garnered input from an international working group of therapeutic VR experts, called the Virtual Reality Committee of Outcomes Research Experts (VR-CORE). This
The document presents the resulting best practice framework informed by expert input, along with specific recommendations for how to conduct high quality VR treatment trials. Although the focus of this document is on VR, the framework also applies to other emerging “XR” technologies, including augmented reality (AR) and mixed reality (MR), as the methodologic considerations for clinical trials are largely similar across XR platforms.

**METHODS**

**Identifying VR-CORE Committee Members**

We performed a systematic review of randomized controlled trials (RCTs) using therapeutic VR to help identify eligible VR-CORE committee members through review of author lists. To garner the largest breadth of studies, the literature search focused on existing meta-analyses of therapeutic VR RCTs identified through PubMed, Google Scholar, and The Cochrane Database of Systematic Reviews using a combination of keywords listed in as follows: (“virtual reality” OR “VR”) AND (“review [pt]” OR “systematic review [pt]” OR “meta-anal*” OR “metaanaly*”). Based on our literature search, and supplemented by recommendations from established experts, we developed a multidisciplinary group to compose VR-CORE, including experts in fields of relevance to developing and testing VR treatments, including user-centered design principles, software design, epidemiology, statistics, and clinical trial methodology. The committee was formulated to balance expertise across clinical disciplines (medicine, pediatrics, surgery, psychology, psychiatry, neuroscience, anesthesia, nursing, rehabilitation) and to reflect multinational perspectives.
Collecting VR-CORE Input

To obtain systematic feedback from the committee, a series of electronic meetings were held to collect and synthesize structured input. An iterative approach was modeled after similar processes we have employed in previous working groups in other fields of healthcare.

Using an online meeting platform that allows users to view and react to each other’s comments (www.Slack.com), committee members initially responded to open-ended “think aloud” prompts (e.g. “When you think about the current state of the clinical VR research, what comes to your mind?”), followed by increasingly specific probes prepared by the moderators (e.g. "What should be the role of human centered design principles in developing VR treatments?"). The full set of questions and responses is listed in Appendix A. Emergent themes and proposed methodologic best practices were culled from the online dialogue, and the resulting recommendations were distributed to the members for synthesis and iterative rephrasing.

RESULTS

Emergent Themes from VR-CORE Meetings

Appendix A provides excerpted transcripts of VR-CORE responses to discussion topics.

Key themes drawn from the online dialogue are summarized in the sections that follow.

Perceptions Regarding Current State of Clinical VR Research

Committee members described the current state of clinical VR research as the "Wild West" with a “lack of clear guidelines and standards.” The state of current VR research was described as “heterogeneous,” often focused “more on the tech rather than the theories behind it.” Committee members expressed concern that much of the current research is “merely...”
"descriptive" in nature, often insufficiently powered, and focused on small case reports, retrospective analyses, and often not employing experimental designs.

**Perceptions Regarding How to Improve VR Literature**

The committee believed it is vital to "include the patients' voice early and often in the development of VR treatments," and that developers must "carefully, systematically, and meticulously seek the patients' feedback" through participatory research and design thinking that involves multidisciplinary collaboration. The committee also called for better definitions and standardization around therapeutic VR study designs.

**Most Important Considerations for Designing and Standardizing Clinical VR Trials**

The committee described various stages for developing and validating VR treatments, beginning with content development in partnership with end-users, progressing through initial clinical testing and safety evaluation, and ending with properly powered RCTs. The committee outlined a wide range of considerations for each stage (**Appendix A**), including the importance of standardizing control groups, selecting clinically relevant outcome measures, reporting which equipment was used in the trial, accounting for dropouts and disqualified participants, and allowing for pragmatic features of each study design, outlined below.

**VR-CORE Clinical Trial Framework**

Although there are fundamental best-practices in study design that apply to all biomedical intervention trials, the committee identified VR-specific attributes that are unique considerations for VR trials. Using the Food and Drug Administration (FDA) Phase I-III pharmacotherapy model as guidance[58], and combining the results of literature synthesis with VR-CORE input, a framework emerged to support three phases of VR clinical study designs,
herein named **VR1, VR2, and VR3**.

VR1 studies focus on content development by working with patient and provider end-users through principles of human centered design. VR2 trials conduct early testing with a focus on feasibility, acceptability, tolerability, and initial clinical efficacy. VR3 trials are RCTs that evaluate clinically important outcomes versus a control condition. Each study should undergo ethical review before initiation. Best practice recommendations for each trial design are described, below.

**VR 1 Studies**

The committee strongly believes that therapeutic VR applications should be designed with direct input from patient and provider end-users.

Lack of patient involvement, poor requirement definitions as well as non-adaptation to user feedback are among the common factors that explain failures of digital interventions[62]. Incorporating patients into the design process enables developers to increase the relevance and effectiveness of VR treatments. The committee stresses that VR treatments should be created with the acknowledgement of the patients' knowledge, attitudes, beliefs, preferences, and expectations of therapeutic VR. VR-CORE refers to a VR1 study as one that results in a VR treatment developed in partnership with patient and provider end-users and follows best practices for patient-centered design.

After their review of literature on human-centered design both generally[63,64] and in relation to digital[62] and VR interventions[65], the committee identified three key principles that are fundamental for developing "desirable, feasible and viable" VR treatments[63]. These
principles, promoting empathy, team collaboration, and continuous user feedback, are detailed below. The committee believe that employing these principles enables development teams to better identify users’ needs, to incorporate user feedback, and institute rapid cycle improvements that generate more relevant products at lower cost. Table 1 outlines the key principles for VR1 studies, described further, below:

**The design process of VR treatments should promote empathy.**

The committee believes that the more attuned a development team is to the specific perspective and needs of patients, the more likely they are to design meaningful VR treatments. Promoting empathy into the design process involves carefully listening to and elucidating patients’ social environment, needs, fears, desires, habits, hopes, aspirations, and expectations. The committee recommends initiating the design process with an *inspiration step, or exercise focused on culling the patients’ voice and understanding their needs, struggles, and experiences.* Table 1 describes best practices for sparking inspiration within the framework of empathy. Different patient profiles and scenarios should be included in this first step. Many techniques can be used to develop empathy and inspiration of the design team. These include qualitative assessments, observations, spending time with users, conducting interviews and user experiments. In addition, a patient journey map can be used to illustrate the interpretation of a story from a patient’s perspective. The working group also recommends seeking input from relevant non-patient end-users, including healthcare providers who may prescribe the VR treatment and/or interact with patient users.
The design process of VR treatments should promote team collaboration. The committee believes that team collaboration is fundamental to collectively designing a VR treatment and synthesizing data collected during the inspiration step. Brainstorming helps generate ideas from the initial corpus of data and findings. Table 1 describes best practices for ideation within the framework of team collaboration. The process of ideation allows team members to think expansively and divergently. As a range of ideas is generated, some ideas will be extreme or ambitious, while others will be achievable. Depending on the time and the available budget, the team decides what ideas should be prototyped further.

The design process of VR treatments should promote continuous user feedback. An effective VR treatment should be developed through continuous user feedback and iterative prototyping, thereby enabling the team to rapidly test their ideas while real-time assessment from end-users. Table 1 describes best practices for VR treatment prototyping within the framework of user feedback. Prototypes should be refined with continuous testing by patient end-users, and failures are viewed as a way for learning and improving the prototype to better meet users’ needs. Hence, the number of defects tends to be lower and less costly in the future solution. To help facilitate the learning process for patients, it is recommended, when feasible, that the research team use a “mirroring” program[67]to allow the research staff to see what the patient is viewing through the VR headset and help them learn the user interface.

In short, the committee believes that the VR1 treatment design process should start with end-users. VR-CORE recommends specifying who the real users are, what they say, see, feel and do. Hence, the implementation of a patient-design approach is an important method to place users at the center of the VR design process. For those researchers that are developing an
open source VR intervention that they would like to share with the academic community for collaborative V1 development process, it is recommended to use a software development platform (e.g. https://github.com/) and cite the latest version of the program within the methods section of VR1 research papers. The committee also recommends the IDEAS checklist developed by Mummah and colleagues as a supplemental, structured guide for conducting a VR1 study[62].

Table 1. Summary of Design principles, Strategies, and Recommended Best Practices for VR1 Studies. See text for details.

<table>
<thead>
<tr>
<th>Design Principles</th>
<th>Strategies</th>
<th>Best Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inspiration through empathizing</td>
<td>Recruitment</td>
<td>Determine the population of interest (who do we need to hear from?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Think about a variety of factors (age, gender, ethnicity, health conditions, social position)</td>
</tr>
<tr>
<td></td>
<td>Observation</td>
<td>Learn about patients and their behaviour by observing them in a clinically relevant context.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observe what patients do in a specific context, what they see, and say.</td>
</tr>
<tr>
<td></td>
<td>Patient interviews</td>
<td>Perform individual cognitive interviews and/or focus groups with patients to learn about their relevant needs, struggles, experiences, fears, aspirations, and expectations.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Document a diverse set of opinions from a variety of patient profiles across age (e.g. above vs. below “digital divide”), co-morbidities, and experience and comfort with technology (e.g. technophiles vs. technophobes).</td>
</tr>
<tr>
<td></td>
<td>Expert interviews</td>
<td>Perform cognitive interviews and/or focus groups with relevant experts representing different points of view, such as treating physicians and other content experts.</td>
</tr>
<tr>
<td></td>
<td>Journey mapping and personas</td>
<td>Define the patient user and describe the sequence of events in which the patient will experience the VR treatment within the context of their illness experience.</td>
</tr>
<tr>
<td>Ideation through team collaboration</td>
<td>Sharing stories and notes</td>
<td>Collect stories, pictures, impressions, and notes about patients’ experiences and behavior.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Share information among team members to generate many ideas through techniques such as storyboarding, story-telling, and mind-mapping.</td>
</tr>
</tbody>
</table>
Generating ideas

Encourage team members to generate ambitious ideas without being judged. The committee believes that idea generation should be distinguished from idea evaluation. After generating ideas, the team evaluates each idea and culls out the most feasible and appropriate for prototyping within technical and budgetary constraints.

Prototyping through continuous user feedback

Building prototype

Convert ideas into tangible figure through drawings or mock-ups and obtain initial user feedback prior to advanced prototyping. Iteratively improve designs with user feedback.

Continuously testing prototype

Test quickly and iterate on the design of the prototype by collecting both positive and negative user feedback. Document all stages of user feedback in resulting VR1 study manuscript.

VR 2 Trials

Once the research team has developed a VR treatment in partnership with end-users, the resulting product should undergo initial assessment among the target patient population within a representative clinical setting, herein termed a VR2 trial. Modeled after the work of Mosadeghi and colleagues[68], the purpose of VR2 trials is to conduct early testing with a focus on acceptability, feasibility, tolerability, and initial clinical efficacy prior to initiating a more definitive VR3 clinical trial, described below. Although developers may opt to bypass a VR2 trial in lieu of a VR3 trial, there is risk in subjecting an incompletely tested intervention to a larger and costlier RCT, and best practices in digital intervention development suggest an intermediary stage between initial VR design and definitive testing[62]. The following sections describe the features of a VR2 trial.
Clinical Setting

In contrast to a VR1 study, which is focused on collaborative content development in a design environment, the VR2 trial evaluates what happens when the VR treatment is placed in the hands of target patients within the intended clinical setting. For example, a VR treatment focused on management of inpatient pain should be tested in an inpatient environment. A VR treatment targeting outpatient stroke rehabilitation should be evaluated in locations where patients receive rehabilitation, such as in a physical therapy center or, if intended, at home. In short, a comprehensive VR2 trial evaluates the VR treatment in the natural setting(s) where the product is intended to be used.

Acceptability

In the context of a VR2 trial, acceptability refers to a patient’s willingness to use the VR treatment. Previous research on therapeutic VR reveals a drop-off between patient eligibility to receive VR and their willingness to try VR[68]. The disconnect emphasizes that many patients are uninterested in using novel health technologies, such as VR, particularly while hospitalized or under duress. Among those who are eligible for a VR trial, some choose not to participate for a wide variety of reasons. Patients may express varying degrees of skepticism, fear, sense of vulnerability, concern regarding psychological consequences, or simply not wanting to be bothered by using the equipment[68]. In a VR2 trial, investigators collect data regarding patient willingness to try the VR treatment, including reasons why they did, or did not, find the intervention to be acceptable for use. Researchers should collect and report acceptability data using techniques such as focus groups, cognitive interviews, or structured questionnaires.
Feasibility

In the context of a VR2 trial, feasibility is the degree to which the VR treatment can be successfully integrated within the flow of usual care. The committee noted that even the best designed VR treatments can face implementation challenges when applied on the front lines of healthcare delivery[68]. It is wise for developers to understand potential barriers early and often, identify workarounds and solutions to these barriers, and only then consider testing their interventions in VR3 RCT trials. For example, patient and providers often seek information regarding the frequency and “dosing” of a VR treatment; these details could be manualized in the context of a VR2 trial. Similarly, treatments deployed in a clinical environment may be unfamiliar to doctors, nurses, and other healthcare providers, allowing researchers an opportunity to study the interaction among staff and proactively identify areas of confusion or misuse. The committee recommends including a table that enumerates patient, provider, technical, and operational barriers to use, identifies root causes, and offers solutions to enhance effectiveness in future clinical applications.

Tolerability

The VR2 trial offers an early opportunity to evaluate patient tolerability of the VR treatment, including both the hardware and software components. Researchers should measure and report the prevalence of patient-reported physical (e.g. vertigo, nausea, “simsickness”) and/or emotional (e.g. fear, anxiety) adverse effects of the VR treatment, along with any discomfort or inconvenience related to the VR equipment (e.g. ill-fitting headset; facial or nasal pain; inability to explore 3D environment fully due to limited mobility, etc.).
Initial Clinical Efficacy

Although the VR2 trial is not designed to test definitively whether a VR treatment is either efficacious or effective, it offers an early opportunity to measure efficacy within the context of a small clinical trial. There is no requirement in a VR2 trial to include a control group, although uncontrolled case series carry a higher risk of bias than controlled studies; even studies with non-randomized concurrent controls, “wait list” controls, or retrospective controls may reduce risk of bias vs. an uncontrolled series.

Regardless of whether there is a control group, investigators should identify a clinically relevant and validated patient reported outcome (PRO) to evaluate for evidence of efficacy. For example, a study evaluating pain might include a standard 11-point numeric rating scale before vs. after exposure to the VR treatment. A study evaluating stroke rehabilitation might measure physical function with the National Institutes of Health (NIH) Patient Reported Outcomes Measurement Information System (PROMIS). Selection of the most appropriate PRO is at the discretion of the research team, but should be carefully justified and must capture the most salient features of patient-reported health intended to improve with the VR treatment.

Table 2. Summary of Best Practice Recommendations for VR2 Trials. See text for details.

<table>
<thead>
<tr>
<th>Trial Attribute</th>
<th>Best Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Population</td>
<td>Study a representative population for whom the VR treatment is intended</td>
</tr>
<tr>
<td></td>
<td>Recruit a large enough sample to represent the breadth and depth of target patients, and to provide statistically stable estimates in descriptive analytics</td>
</tr>
<tr>
<td>Clinical Setting</td>
<td>Select a clinical setting that represents the intended environment for the VR treatment to be used (e.g. inpatient vs. outpatient; clinic vs. home-based, etc.)</td>
</tr>
<tr>
<td>Assessment of</td>
<td>Collect data regarding patient willingness to try the VR treatment, including reasons why</td>
</tr>
</tbody>
</table>
Acceptability

they did, or did not, find the intervention to be acceptable for use. Researchers should collect and report acceptability data using techniques such as focus groups, cognitive interviews, or structured questionnaires.

Assessment of Feasibility

Conduct patient and provider interviews to identify potential barriers and facilitators, to using the VR treatment in the intended clinical environment
Collect information regarding the optimal frequency and "dosing" of a VR treatment; consider manualizing these details where possible.
Study interactions among staff and proactively identify areas of confusion or misuse.
Consider including a table that enumerates patient, provider, technical, and operational barriers to use, identifies root causes, and offers solutions to enhance effectiveness in future clinical applications.

Assessment of Tolerability

Measure and report the prevalence of patient-reported physical and/or emotional adverse effects of the VR treatment, along with any discomfort or inconvenience related to the VR equipment

Assessment of Initial Clinical Efficacy

Identify and justify selection of a clinically relevant and validated patient reported outcome (PRO) to evaluate for evidence of efficacy.
Measure the PRO before vs. after receipt of the VR treatment in sufficiently powered cohort; consider comparing results against non-randomized concurrent or retrospective control groups, where available

VR3 Trials

The most definitive clinical validation of a VR treatment is the VR3 trial, a prospective, adequately powered, methodologically rigorous RCT evaluating clinical outcomes and safety in target patients receiving the VR treatment vs. a control condition. Although therapeutic mechanism of action may be studied as secondary goal in a VR3 trial (e.g. through neuroimaging, blood biomarkers, physiologic testing, etc.), the principal goal is to evaluate the treatment's impact on a clinically meaningful patient outcome rather than surrogate markers.

Although the committee acknowledged understandable cost and resource barriers to conducting VR3 trials, there was broad agreement that RCTs are of equal scientific importance in therapeutic VR as they are to any other form of treatment and should be prioritized

289 VR3 Trials

290 The most definitive clinical validation of a VR treatment is the VR3 trial, a prospective, adequately powered, methodologically rigorous RCT evaluating clinical outcomes and safety in target patients receiving the VR treatment vs. a control condition. Although therapeutic mechanism of action may be studied as secondary goal in a VR3 trial (e.g. through neuroimaging, blood biomarkers, physiologic testing, etc.), the principal goal is to evaluate the treatment’s impact on a clinically meaningful patient outcome rather than surrogate markers.

296 Although the committee acknowledged understandable cost and resource barriers to conducting VR3 trials, there was broad agreement that RCTs are of equal scientific importance in therapeutic VR as they are to any other form of treatment and should be prioritized
Whenever possible, multicenter collaborations may facilitate VR3 trials by combining patients and resources through shared protocols. The features of a VR3 trial are described, below.

**Standardization of Intervention and Patient Population**

Having been developed in a VR1 study and initially tested in a VR2 trial, the study intervention should be clearly described in preparation for a VR3 trial. Researchers should provide details regarding the equipment used, visualizations employed (with representative screenshots or videos), frequency, duration, and timing of use. Optimally, the intervention should be manualized, but at the very least enough details should be provided to allow other investigators to repeat the trial, if so desired. The TIDIER checklist provides a useful framework for describing study interventions[71] and should be applied to VR treatments. The target patient population should be clearly described, including explicit inclusion and exclusion criteria employed. Certain exclusion criteria may be standardized among VR trials, such as a history of significant motion sickness, active nausea and vomiting, or epilepsy.

**Selection of Control Condition**

The committee acknowledged that there is no perfect or standardized control condition for all VR treatment trials; the optimal control depends on the patient population, proposed mechanism of action of the intervention, and clinical setting, among other considerations. Selection of the control is at the discretion of the research team but should be justified and explained. The committee described a hierarchy of control conditions, ranging from “usual care” without any active intervention, to passive visualizations on a two-dimensional screen, to non-immersive visualizations within a headset, to immersive but passive experiences within a headset, to immersive and active experiences within a headset. Selecting the optimal control
may be guided by considering the hypothesized target of engagement and the proposed mechanism of action.

**Randomization**

Randomization should be described and ideally achieved using an appropriate computer program (e.g. MS Excel Random Number Generator) or random number tables without involvement of the investigators who enrolled the patients. Longitudinal studies with repeated measures may consider using micro-randomization, where individuals are randomized to treatment at multiple time points throughout the study, thereby acting as their own control.

**Blinding and Concealment of Allocation**

The committee acknowledge that blinding and concealment can be challenging but identified techniques to incorporate these RCT principles within the constraints of VR research. For example, Spiegel and colleagues achieved concealment of allocation in an RCT comparing a library of VR content vs. a “health and wellness” television channel in hospitalized patients experiencing pain. At the time of consent, the researchers explained to patients that the study was comparing “two different audiovisual experiences designed to reduce pain”, but did not describe the details of the competing interventions. Patients randomized to the television intervention did not know that VR was the other condition, and vice versa. This approach may reduce the “novelty effect” of receiving VR rather than a familiar experience like television.

Equipoise may also be achieved by exposing patients in both arms to headsets, but varying the content viewed within the headset (e.g. immersive vs. non-immersive; active vs. passive). At a minimum, study analysts should be blinded to patient group allocation, allowing for an unbiased evaluation of the data without knowledge of study group. Patients should be asked not to reveal...
details of the program they experienced to decrease chance of unblinding the study analysts.

The measurement of perceived group assignment at the end of the study can help assess the success of blinding within the study. This should be done at the discretion of the research team.

Endpoints

Like the VR2 trial, VR3 trials must pre-specify a clinically relevant and validated PRO as the primary endpoint. The study must be appropriately powered to demonstrate a minimally clinically important difference (MCID)\[74\] in that endpoint between the VR treatment and control arms. The psychometrics of PRO measurement are beyond the scope of this document, but existing references may assist investigators in protocol development\[74,75\]. Secondary endpoints may include a variety of clinical, imaging, biometric, and/or physiologic surrogate markers, as deemed appropriate by the study team. Like VR2 trials, potential adverse events must be prospectively measured and reported.

Study Duration

VR3 studies should monitor patients for a sufficient period to determine whether the VR treatment meaningfully impacts clinically important outcomes. One-time, short-term evaluations may be insufficient to evaluate the true clinical value of an intervention. Follow-up over several days may be appropriate if only focusing on a hospital stay, but measurement over weeks, or even months, may be necessary to assess the impact on long-term clinical benefits.
VR-CORE suggests that the primary outcome be reported as the before vs. after difference in difference between study arms, with accompanying 95% confidence intervals. For example, the change in mean PRO score before vs. after the VR intervention should be compared against the change in mean PRO score before vs. after the control intervention. In addition, the panel recommends pre-defining a binary response criterion, guided by the MCID of the primary endpoint. The proportion achieving the MCID should be reported and compared between groups, and the resulting number needed to treat (NNT) should be calculated.

The primary analyses should use the intention-to-treat (ITT) population, including all patients randomized regardless of follow-up or receipt of study interventions. However, per-protocol (PP) analysis may be appropriate in certain situation, such as if patients refuse the VR treatment after randomization; in this instance, reporting the rate of refusal would be important, but investigators might also seek to compare therapeutic responses only among those receiving the intervention.

Multivariable analysis may be useful in adjusting for pre-specified confounding factors (especially if not equally distributed in the study groups) and to explore independent predictors of outcomes. To perform a multivariable analysis, it is optimal to have at least 10 (and preferably 20) observations for each independent variable included in the multivariable model. Given adequate repeated measures, multilevel modeling may be used to examine the extent to which the intervention acts on person-level (i.e., between-subject) or momentary (i.e., within-subject) factors.
**Reporting the Trial**

VR3 trials must be registered in a publicly accessible registry (e.g., www.clinicaltrials.gov). All completed trials should be published, whether positive or negative. The Consolidated Standards for Reporting Trials (CONSORT) guidelines provide the framework for reporting RCTs[77], and they should be followed in VR3 trials. VR3 trials must include a CONSORT diagram demonstrating the flow of patients through each stage of the trial, including the number screened to the number randomized into each study group to the number analyzed.

**Table 3. Summary of Best Practice Recommendations for VR3 Trials.** See text for details.

<table>
<thead>
<tr>
<th>Trial Attribute</th>
<th>Best Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Population</td>
<td>Study a representative population for whom the VR treatment is intended</td>
</tr>
<tr>
<td></td>
<td>The target patient population should be clearly described, including explicit inclusion and exclusion criteria employed.</td>
</tr>
<tr>
<td>Clinical Setting</td>
<td>Select a clinical setting that represents the intended environment for the VR treatment to be used (e.g. inpatient vs. outpatient; clinic vs. home-based, etc)</td>
</tr>
<tr>
<td>Standardizing Intervention</td>
<td>Provide details regarding the equipment used, visualizations employed, frequency, duration, and timing of use for VR treatment.</td>
</tr>
<tr>
<td></td>
<td>Consider following TIDIER checklist[71] as a useful framework for describing features of the VR treatment</td>
</tr>
<tr>
<td>Selecting Control Condition</td>
<td>Selecting and justify the control condition(s) by considering the hypothesized target of engagement and the proposed mechanism of action.</td>
</tr>
<tr>
<td>Randomization</td>
<td>Randomization should be achieved using an appropriate computer program (e.g. MS Excel Random Number Generator) or random number tables without involvement of the investigators who enrolled the patients.</td>
</tr>
<tr>
<td>Blinding and Concealment of Allocation</td>
<td>Describe efforts to conceal allocation of the study intervention to the participants</td>
</tr>
<tr>
<td></td>
<td>Describe efforts to blind patient, providers, and analysts, wherever possible</td>
</tr>
<tr>
<td></td>
<td>Measure perceived group assignment to assess success of blinding</td>
</tr>
</tbody>
</table>
| Endpoints | Pre-specify a clinically relevant and validated patient reported outcome (PRO) as the primary endpoint.  
Trial must be appropriately powered to demonstrate a minimally clinically important difference (MCID)[70] in primary endpoint between the VR treatment and control arms.  
Secondary endpoints may include a variety of clinical, imaging, biometric, and/or physiologic surrogate markers, as deemed appropriate by the study team.  
Potential adverse events must be prospectively measured and reported. |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Duration</td>
<td>Select and justify follow-up period that is sufficient to determine whether the VR treatment meaningfully impacts clinically important outcomes.</td>
</tr>
</tbody>
</table>
| Presentation and Analysis of Results | Report the before vs. after difference in difference in the primary outcome measure between study arms, with accompanying 95% confidence intervals.  
Pre-define a binary response criterion, guided by the minimal clinically important difference (MCID) of the primary endpoint. The proportion achieving the MCID should be reported and compared between groups, and resulting number needed to treat (NNT) calculated.  
Use intention to treat (ITT) analysis for primary outcome assessment.  
Per protocol analysis may be reported if pre-specified as relevant.  
To perform a multivariable analysis, it is optimal to have at least 10 (and preferably 20) observations for each independent variable included in the multivariable model. |
| Reporting the Trial | Trial must be registered on a publicly accessible registry (e.g., www.clinicaltrials.gov).  
All completed trials should be published, whether positive or negative.  
The Consolidated Standards for Reporting Trials (CONSORT) guidelines provide the framework for reporting RCTs[77], and they should be followed in VR3 trials.  
Include a CONSORT diagram demonstrating the flow of patients through each stage of the trial, including the number screened to the number randomized into each study group to the number analyzed. |

**CONCLUSION**

To improve methodological quality in the therapeutic VR literature, the VR-CORE international working group presents a three-part framework for best practices in developing and testing VR treatments. This framework may be used to facilitate development of high quality, effective, and safe VR treatments that meaningfully improve patient outcomes. Patients,
providers, payers and regulators may consider this framework when assessing the validity of VR treatments.
In addition to the authorship team, the following committee members actively participated in the online discussion forums and are acknowledged here as contributors to this document:

- Tom Caruso, MD, M.Ed – Stanford University, Palo Alto, USA
- Ali Fardinpour, PhD – University of Western Australia, Perth, Australia
- Diane Gromala, PhD – Simon Fraser University, Vancouver, Canada
- Rafael Grossman, MD – Eastern Maine Healthcare System, Bangor, USA
- Kate Hardy, Clin. Psych.D – Stanford, Palo Alto, USA
- Constance Johnson, Ph.D. – Duke, Durham, USA
- Ted Jones, PhD – Pain Consultants of East Tennessee, Knoxville, USA
- Kate Laver, Ph.D. – Flinders University, Adelaide, Australia
- Sheila Parinas, BSN RN – University of Illinois, Champaign, USA
- Les Posen, PhD – Flightwise, Murrumbeena, Australia
- Barbara Rothbaum, Ph.D., ABPP – Emory, Atlanta, USA
- David Thomas, Ph.D. – National Institute of Drug Abuse (NIDA), Bethesda, USA
- Stefan Van Rooijen, M.D. Ph.D. – Maastricht University, Limburg,
- Herve Rosay, M.D. – Centre Leon Berard, Lyon, France
- Earl Scott, M.D. – HCA Midwest Physician, Kansas City, Missouri, USA
- Andrea Won Stevenson, Ph.D. – Columbia, Ithaca, USA
Competing interest

Authors:

Dr. Rizzo’s research relating to this paper has been funded by the NIH, NSF, US Army Research Office, TATRC, USAMRAA, Department of Veterans Affairs, Kesseler Foundation. All other authors (BB, SB, CK, XL, SC, ID, KB) have no conflicts of interest.

VR-CORE members:

Dr. Rothbaum owns equity in Virtual Better, Inc., which is developing products related to virtual reality researcher related to this paper. The terms of this arrangement have been approved by Emory University in accordance with its conflict of interest policies. Dr. Johnson receives funding through the NIH to study virtual environments. Some of Dr. van Rooijen VR research has been funded by Phillips, Inc. All other members of the VR-CORE (TC, AF, DG, RG, KH, TJ, KL, SP, LP, DT, HR, ES, AWS) have no conflicts of interest.
Funding: No funding was used in the creation of the project

Multimedia Appendix A: [Excerpted transcripts of VR-CORE responses to selected discussion topics. Key themes and phraseology included in manuscript are highlighted. Note that not all committee members responded to all questions]
REFERENCES


Norman GR, Sloan JA, Wyrwich KW. Interpretation of Changes in Health-related Quality of Life. Med Care 2003 May;41(5):582–592. PMID:12719681
