



**HAL**  
open science

# Reporting Guidelines for the Early-Phase Clinical Evaluation of Applications Using Extended Reality: RATE-XR Qualitative Study Guideline

Johan H Vlake, Denzel L Q Drop, Jasper van Bommel, Giuseppe Riva, Brenda K Wiederhold, Pietro Cipresso, Albert S Rizzo, Barbara O Rothbaum, Cristina Botella, Lotty Hooft, et al.

## ► To cite this version:

Johan H Vlake, Denzel L Q Drop, Jasper van Bommel, Giuseppe Riva, Brenda K Wiederhold, et al.. Reporting Guidelines for the Early-Phase Clinical Evaluation of Applications Using Extended Reality: RATE-XR Qualitative Study Guideline. *Journal of Medical Internet Research*, 2024, 26, 10.2196/56790 . hal-04841792

**HAL Id: hal-04841792**

**<https://hal.science/hal-04841792v1>**

Submitted on 16 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Tutorial

# Reporting Guidelines for the Early-Phase Clinical Evaluation of Applications Using Extended Reality: RATE-XR Qualitative Study Guideline

Johan H Vlake<sup>1,2</sup>, BSc; Denzel L Q Drop<sup>1</sup>, MD; Jasper Van Bommel<sup>1</sup>, MD, PhD; Giuseppe Riva<sup>3,4</sup>, PhD; Brenda K Wiederhold<sup>5</sup>, BCN, BCB, MBA, PhD; Pietro Cipresso<sup>3,6</sup>, PhD; Albert S Rizzo<sup>7</sup>, PhD; Barbara O Rothbaum<sup>8</sup>, PhD; Cristina Botella<sup>9,10</sup>, PhD; Lotty Hoof<sup>11</sup>, PhD; Oscar J Bienvenu<sup>12</sup>, MD, PhD; Christian Jung<sup>13,14</sup>, MD, PhD; Bart Geerts<sup>15</sup>, MSc, MBA, MD, PhD; Evert-Jan Wils<sup>2</sup>, MD, PhD; Diederik Gommers<sup>1</sup>, MD, PhD; Michel E van Genderen<sup>1</sup>, MD, PhD; RATE-XR Expert Group<sup>16</sup>

<sup>1</sup>Department of Intensive Care, Erasmus Medical Center, Rotterdam, Netherlands

<sup>2</sup>Department of Intensive Care, Franciscus Gasthuis & Vlietland, Rotterdam, Netherlands

<sup>3</sup>Applied Technology for Neuro-Psychology Lab, IRCCS Istituto Auxologico Italiano, Milan, Italy

<sup>4</sup>Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy

<sup>5</sup>Virtual Reality Medical Center, San Diego, CA, United States

<sup>6</sup>Department of Psychology, University of Turin, Turin, Italy

<sup>7</sup>Medical Virtual Reality Lab, University of Southern California Institute for Creative Technologies, Los Angeles, CA, United States

<sup>8</sup>Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, United States

<sup>9</sup>Department of Basic Psychology, Clinic, and Psychobiology, University Jaume I, Castellón, Spain

<sup>10</sup>CIBER de Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, Madrid, Spain

<sup>11</sup>Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

<sup>12</sup>Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, United States

<sup>13</sup>Department of Cardiology, Pulmonology, and Vascular Medicine, Medical Faculty, University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

<sup>14</sup>Cardiovascular Research Institute Düsseldorf (CARID), Medical Faculty, University Hospital of Düsseldorf, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

<sup>15</sup>R&D BV, Healthplus.ai, Amsterdam, Netherlands

<sup>16</sup>See Acknowledgments

**Corresponding Author:**

Michel E van Genderen, MD, PhD

Department of Intensive Care

Erasmus Medical Center

Doctor Molewaterplein

Rotterdam, 3015 GD

Netherlands

Phone: 31 107040704

Email: [m.vangenderen@erasmusmc.nl](mailto:m.vangenderen@erasmusmc.nl)

## Abstract

**Background:** Extended reality (XR), encompassing technologies such as virtual reality, augmented reality, and mixed reality, has rapidly gained prominence in health care. However, existing XR research often lacks rigor, proper controls, and standardization.

**Objective:** To address this and to enhance the transparency and quality of reporting in early-phase clinical evaluations of XR applications, we present the “Reporting for the early-phase clinical evaluation of applications using extended reality” (RATE-XR) guideline.

**Methods:** We conducted a 2-round modified Delphi process involving experts from diverse stakeholder categories, and the RATE-XR is therefore the result of a consensus-based, multistakeholder effort.

**Results:** The guideline comprises 17 XR-specific (composed of 18 subitems) and 14 generic reporting items, each with a complementary Explanation & Elaboration section.

**Conclusions:** The items encompass critical aspects of XR research, from clinical utility and safety to human factors and ethics. By offering a comprehensive checklist for reporting, the RATE-XR guideline facilitates robust assessment and replication of early-stage clinical XR studies. It underscores the need for transparency, patient-centeredness, and balanced evaluation of the applications of XR in health care. By providing an actionable checklist of minimal reporting items, this guideline will facilitate the responsible development and integration of XR technologies into health care and related fields.

(*J Med Internet Res* 2024;26:e56790) doi: [10.2196/56790](https://doi.org/10.2196/56790)

## KEYWORDS

extended reality; XR; virtual reality; augmented reality; mixed reality; reporting guideline; Delphi process; consensus; computer-generated simulation; simulation; virtual world; simulation experience; clinical evaluation

## Introduction

Extended reality (XR) encompasses various forms of computer-generated reality, including augmented reality (AR), mixed reality (MR), and virtual reality (VR). XR, mainly in the form of VR, has rapidly emerged in health care, particularly in fields such as mental health, intensive care medicine, surgery, pain management, and rehabilitation [1-4]. Much like other transformative technologies as artificial intelligence (AI) algorithms, the field of XR has witnessed an exponential surge in research and applications: from 1992 to 2005, merely up to 100 publications were recorded yearly, but this number has steadily increased, with over 1000 publications annually since 2018 [2]. Notably, the US Food and Drug Administration has been approving a growing number of XR-based devices, underscoring its escalating clinical significance [5].

Despite this expanding landscape of XR research in health care, most studies primarily focus on treatment effects, tend to be small and heterogeneous, and often lack proper control conditions [6]. Consequently, comparing XR studies is challenging, as scientific rigor is often lacking and they pose several unique implementation and technological challenges in health care [7-9]. To overcome these challenges and to promote optimal reporting of XR-based interventions' clinical utility, a more structured approach is essential. This approach should encompass technological, methodological, and safety aspects to support a more objective understanding of the validity and

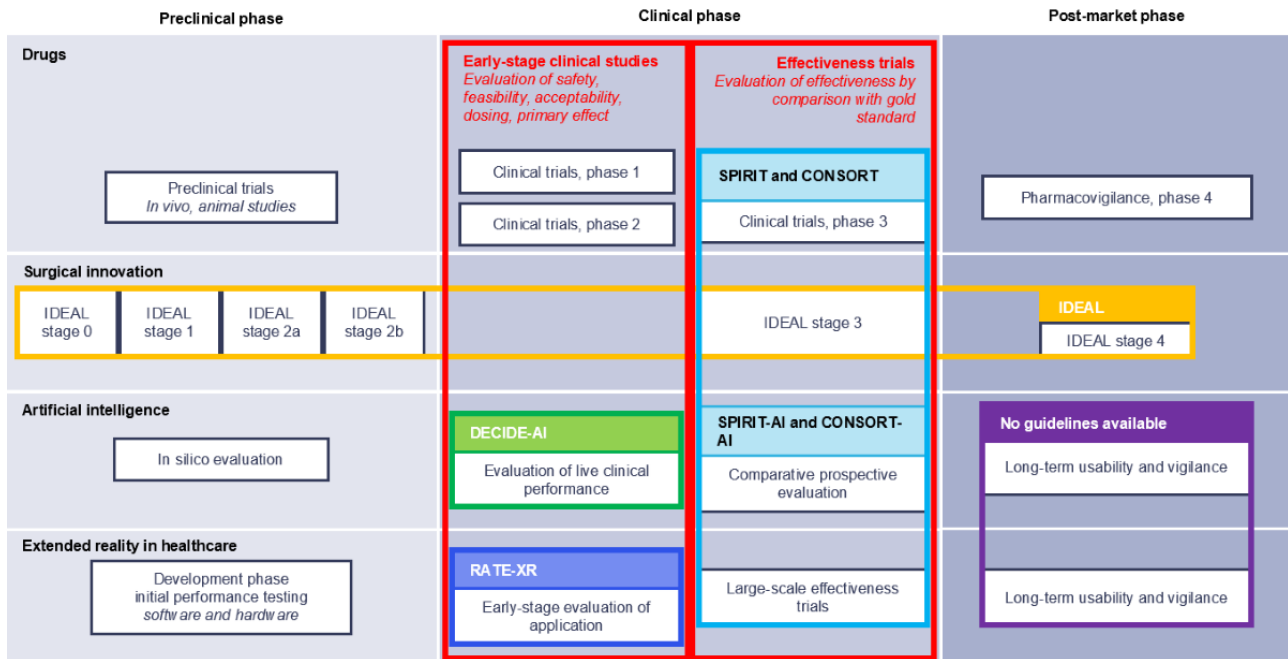
generalizability of findings [10]. The challenges of early-stage clinical evaluation of applications using XR (Textbox 1) share similarities to those of other innovative technologies and interventions such as developing and implementing surgical innovations or AI models [11-14].

Early-stage clinical evaluation of XR systems plays a pivotal role in bridging the gap between preclinical technological development and large-scale effectiveness trials. Existing guidelines such as the VR core model stage 2/3 recommendations, SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) and CONSORT (Consolidated Standards of Reporting Trials) statements, including their AI extensions, and the IDEAL guidelines offer valuable insights into the design and reporting of clinical trials (Figure 1) [12,13,15,16]. However, they generally focus on later stages of clinical research or do not adequately address the specific challenges of developing and evaluating XR technologies in early clinical settings. These include the rapid evolution of XR hardware and software and the specific safety, usability, and ethical considerations that arise in these contexts. To address these gaps, the reporting for the early-phase clinical evaluation of applications using extended reality (RATE-XR) guideline specifically tailors its recommendations to support transparent reporting and effective evaluation of XR applications from the developmental phase through to early clinical trials, ensuring that these innovations can be safely and effectively integrated into health care practice.

**Textbox 1.** The challenges of early-phase clinical evaluation of applications using extended reality (XR). The clinical evaluation of applications using XR presents several challenges, all of which will likely be encountered at early stages. This textbox represents several of these challenges.

- Allow for a continuous changing nature of the software applications using XR and its hardware (due to early prototyping and version updates)
- Account for technical errors negatively impacting the reliability and consistency of clinical evaluations
- Evaluate the generalizability of findings across sites and populations
- Deal with ethical considerations in this early stage of research
- Deal with a variety of clinical trial endpoints for applications due to the wide range of intended uses
- Account for user variability and consequently the arising bias as users may not be trained adequately or may not be familiar with XR technology
- Incorporate XR applications into the standard workflow
- Absence of established methodologies and frameworks
- Having sufficient image quality of XR devices for users, due to their diversity and constantly evolving technological characteristics
- Create a usability profile applicable to various working environments

**Figure 1.** Comparison of development pathways for drug therapies, surgical innovation, artificial intelligence, and extended reality in health care. The colored lines represent reporting guidelines, some of which are study design–specific (SPIRIT or CONSORT and SPIRIT or CONSORT-AI); others are stage-specific (IDEAL and RATE-XR). Depending on the context, more than one study design can be appropriate for each stage. CONSORT: Consolidated Standards of Reporting Trials; RATE-XR: reporting for the early-phase clinical evaluation of applications using extended reality; SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials.



Early-stage clinical evaluation of XR interventions must prioritize clinical utility, safety, and human factors challenges in real-life clinical settings. Factors such as cybersickness, customizability and duration of the XR content, treatment frequency, and immersiveness contribute to safety and feasibility considerations and must be addressed transparently. A one-size-fits-all model is often not feasible, and neglecting safety profiles and rushing into large-scale trials can jeopardize patient well-being, which is ethically unacceptable. In terms of ethics, patient-centeredness and commercial interests must be addressed in the early stages of XR development. Currently, studies too often prioritize assessing commercial products but fail to address essential clinical research elements or tailoring content for medical applications, while understanding the interaction between XR technology and human factors is essential [17-19]. The clinical context should be the starting point in the development of medical XR systems, involving patients and health care providers as primary stakeholders early in the process to design systems that optimally address their needs, beyond placing an initial focus on commercial product evaluation. Moreover, variations in hardware, software, and content selection are complex to assess and often underreported [20].

To address these challenges and with the aim of improving the consistency, safety, knowledge generation, and applicability of XR research in the health care domain, we undertook a robust 2-phase modified Delphi process. This collaborative effort engaged strong and diverse stakeholder engagement and resulted in the development of the RATE-XR guideline. Here, we present the development process, key recommendations, and their implications for the XR health care field.

## Methods

### RATE-XR Guideline Development

The RATE-XR guideline was developed through an international expert consensus process adhering to the EQUATOR Network’s recommendations for guideline development [21].

### Establishment of the Steering Committee

To guide the development of the RATE-XR guideline, a Steering Committee was assembled, detailed in Table S1 in [Multimedia Appendix 1](#). The committee was selected by the project initiators—JHV, JvB, and MEvG—to ensure a diverse representation of expertise within the XR and research domains. Members of our study group, including CJ, DLQD, DG, EJW, and OJB, were included for their direct contributions to this project. Additionally, we engaged 5 of the most influential authors in XR research as cited in a recent JMIR article—GR, BKW, PC, ASR, and CB—to incorporate a broad range of perspectives and expertise [2]. We also invited BOR, a pioneer in clinical VR, and BJB, known for his work in VR trial design and implementation. Finally, to incorporate expertise in methodology and guideline development, LH and BG, both experienced in developing previous reporting guidelines, were included [13]. We conducted a modified Delphi process based on previous guideline development consisting of 2 rounds of feedback from participating experts followed by virtual consensus meetings and qualitative evaluation by an independent evaluation committee [13,22,23].

### Ethical Considerations

The project was approved by the Medical Ethics Committee of the Erasmus Medical Centre, Rotterdam (approval 2022-0623) and registered with the EQUATOR (Enhancing the Quality and

Transparency of Health Research) network. Informed consent was obtained from all members of the Steering Committee, all participants of the Delphi rounds, and all members of the evaluation committee.

### Generation of the Initial Item List

An initial list of 61 candidate items (with subitems) was composed by 2 authors (JV and MEvG) and was based on (1) scientific reports on trials examining XR-based studies in health care [24-27], (2) recently published innovative technology guidelines [13,28], (3) methodological and evaluative challenges concerning the application of XR in health care [14,16], (4) a Cochrane Systematic Review on the clinical use of XR [29], and (5) institutional documents [30-32]. Hereafter, the candidate item list was commented on by the Steering Group members (Steering Group Round).

### Expert Recruitment

Experts were recruited using five distinct approaches: (1) invitations to experts that were endorsed by the Steering Group members, (2) invitations to authors of publications identified through the preliminary literature search, (3) a call for contributions published within a medical journal [33], (4) consideration of professionals proactively reaching out to the Steering Group, and (5) invitations to experts recommended by the Delphi participants (snowballing). Prior to the initiation of recruitment, 17 target stakeholder groups were defined: clinicians, engineers or computer scientists, methodologists, statisticians, implementation specialists, entrepreneurs, epidemiologists, journal editors, allied health professionals, policy makers or official institutional staff, administrators or hospital management, researchers, ethicists, private sector representatives, patient representatives, funders, and psychologists or psychiatrists.

### The Delphi Process

The Delphi process consisted of 2 rounds, and the Delphi surveys were designed and distributed using the Castor Electronic Data Capture web application (Castor EDC). The first round encompassed 2 parts. In the first part, participants answered 6 open-ended inquiries that address facets considered essential to be reported on during early-phase clinical evaluation. In the second part, Delphi participants were tasked with rating, from 1 to 9, the significance of items in the initial list. Ratings of 1 to 3 indicated insignificance, 4 to 6 denoted importance without being pivotal, and 7 to 9 implied that items were both important and critical. In addition to rating the items, participants were prompted to offer commentary on items and propose new additions. Thematic analysis of the open-ended questions was independently conducted by 2 Steering Group members (JV and MEvG), with any disagreements being resolved through consensus. Identified themes were used to determine whether any important themes were missing in the item list, along with newly proposed items to complement the item list. Hereafter, a summary score, including the median, 25th percentile, 75th percentile, mean, SD, proportion of participants scoring the item above 7 or below 3, and stakeholder groups with a median

of  $\leq 2$  or  $\geq 2$  points from the overall median were calculated for each item. Prespecified inclusion cutoffs were determined as an item scoring a mean  $\geq 7$  and exclusion as an item scoring a mean  $\leq 3$ . Based on these results, a revised item list for the second Delphi round was generated.

In the second Delphi round, participants were presented with the outcomes of the first round, along with the revised item list. Participants were tasked with reevaluating all items in a manner akin to the first Delphi round and were invited to comment on content and wording. Both Delphi round surveys and outcomes are accessible through the Open Science Framework (OSF) [34]. All analyses were performed using R for Statistics (R Foundation for Statistical Computing).

### Consensus Meeting

Virtual consensus meetings were held on 3 separate occasions between June 12 and 15, 2023, with the aim of finalizing content and refining the phrasing of items within the RATE-XR reporting guideline. To ensure a balanced representation of key stakeholders throughout the XR field and geographic diversity, we engaged 18 experts with diverse expertise and backgrounds (Tables S2-S4 in [Multimedia Appendix 1](#)). Throughout the consensus meetings, all items from the second Delphi round were subject to discussion and anonymous voting, facilitated by the Mentimeter platform [35]. The voting process was overseen by a chairman and observed by a designated observer. For an item to be ultimately included in the definitive guideline, a predefined threshold of 80% agreement among Consensus Group members was necessary, excluding abstentions and blank votes.

### Qualitative Evaluation

After finalizing both the guideline and the Explanation & Elaboration note, a qualitative evaluation was conducted by a panel of 14 experts (Note 1 in [Multimedia Appendix 1](#)) possessing significant experience in implementing or peer-reviewing literature relevant to applications using XR. None of the experts involved in the qualitative evaluation were affiliated with the Consensus Group. Their input focused on evaluating the clarity and usability of each XR-specific item using a custom form, which is available on the OSF platform [34]. In total, 3 reviewers (JV, DD, and MEvG) independently reviewed the provided comments to assess the necessity for revisions in the wording of items or their corresponding Explanation & Elaboration sections. The review process was structured to ensure comprehensive coverage and unbiased analysis of the feedback. Disagreements between reviewers were resolved through consensus, ensuring a balanced interpretation of the qualitative data. This methodological rigor enhances the reliability of the modifications made to the RATE-XR guidelines based on stakeholder feedback. To enhance comprehension of key concepts within the guideline, a glossary of terms ([Table 1](#)) was composed. All Consensus Group members approved the modifications, the final guideline, and the complementary Explanation & Elaboration note.

**Table 1.** Glossary of terms<sup>a</sup>.

Terms	Explanations
Application	The software, program, intervention, or modality using an XR device or hardware.
Artificial intelligence algorithm	"Science of developing computer systems which can perform tasks normally requiring human intelligence" based on a mathematical model responsible for learning from data and producing an output [36].
Augmented reality	A technology that overlays digital information onto the real-world environment, viewed through an augmented reality headset or glasses in order to enhance the user's perception of reality. Augmented reality is part of the extended reality (XR) technologies.
Bias	Systematic difference in treatment of certain objects, people, or groups in comparison to others [37].
Clinical utility	The practical value and usefulness of the application using XR.
Commercial name of application	Trademarked or branded name under which the specific software application or hardware device is sold.
Commercial product	An item originally manufactured for sale, lease, or license to the general public.
Cybersickness	A form of motion sickness or discomfort experienced by individuals while using XR devices.
Early-phase studies	Studies in the initial stages of investigation where applications using XR devices are tested and evaluated, focusing on safety, dosage, feasibility, and potential efficacy of the intervention involving a relatively small number of participants.
XR	An umbrella term to encompass the spectrum of immersive technologies consisting of virtual reality, augmented reality, and mixed reality.
Flow diagram	A visual representation using symbols and arrows to illustrate the sequential steps, processes, or decisions within the workflow or procedure.
Hardware	The physical components and equipment that make up the XR system.
Human factors	Also called ergonomics. "The scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and methods to design in order to optimize human well-being and overall system performance" (International Ergonomics Association).
Immersiveness	The degree to which the experience captivates and engages the user, involving a deep sense of presence and absorption within a simulated environment.
Immersive virtual reality	A computer-generated, 3D artificial environment using a head-mounted display and therefore completely surrounding the user's senses. This way the user is brought from the real world into the artificial, virtual world.
Mixed reality	A technology combining elements of both virtual reality and augmented reality. It integrates digital content and virtual objects into the real-world environment, allowing users to interact with and manipulate these virtual elements as if they were part of their physical surroundings.
Performance	How well the application, device, system, or technology functions and executes its intended task or functionalities.
Preclinical	Pertaining to the phase of research prior to clinical trials targeting actual patients.
Prespecified outcomes	Specific defined results, goals, or expectations that are determined and established in advance, prior to performing the study.
Real clinical setting	Pertaining to the observation and treatment of actual patients, instead of preclinical users or simulated scenarios.
Software	The applications, programs, and digital content are specifically designed to interact with the XR system.
Usability	"Extend to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [38].

<sup>a</sup>The definitions provided apply to the specific context of RATE-XR and the use of the terms in the guideline. They are not necessarily generally accepted definitions and may not always be entirely suitable for other research domains.

## Results

### Initial Item List

Based on the yielded 97 comments and 22 proposals for new items, 18 items were appended and 29 were subjected to reorganization (through merging or splitting, resulting in 16 items). Additionally, wording was amended, and items were categorized into XR-specific and general items. The final initial item list resulted in 71 Delphi items, subdivided into 41

XR-specific and 30 general reporting items, 6 categories, and 22 subcategories, and was approved on by all Steering Group members (see digital file in the OSF platform) [34].

### Delphi Rounds

A total of 124 individuals expressed their interest and completed the participation form for the first Delphi round, of whom 22 were unqualified due to lack of XR-related experience. Among the 102 experts who received the first Delphi questionnaire, 93

(91%) completed the questionnaire. The participants included 13 Steering Group members, 38 identified from Steering Group recommendations, 13 from proactive contacts or correspondence, and 29 through snowballing. In total, 112 experts were invited to participate in the second Delphi round, of which 96 (86%) responded. In total, 82 of these experts also participated in the first Delphi round (continuity rate: 88%). Collectively, the participating experts represented 14 countries, and all stakeholders were represented (Supplementary Note 1 and Tables S5-S8 in [Multimedia Appendix 1](#)).

The first Delphi round yielded over 17,300 words of unstructured text to the open-ended inquiries, along with 6603 item scores, 256 item comments, and 97 newly proposed items. Thematic analysis identified 146 themes, of which 88 were covered in existing items, 22 were integrated into or added to the provisory Explanation & Elaboration note, 28 were used to amend existing items, 2 were selected as new items, and 6 were dropped as they were determined to be outside of the reporting guideline scope. Eventually, 5 items remained unchanged, 27 items were amended or rephrased, 36 items were merged or split into 14 items, 3 items were dropped, and 5 items were added (Figures S1 and S2 in [Multimedia Appendix 1](#)). The 3 items that were dropped were related to production costs of the XR module and were dropped due to low consensus in the

scoring exercise and congruent comments that these items were out of scope. The revised item list eventually comprised 51 Delphi items in 45 reporting items, subdivided into 22 XR-specific and 23 general reporting items. The second Delphi round yielded 4896 item scores and 372 comments.

### Consensus Meeting

In total, 32 items received endorsement for integration into the RATE-XR guideline during the consensus meetings—17 items specific to XR and 14 encompassing general reporting. A summary of the Consensus Meetings votes is presented in Table S9 in [Multimedia Appendix 1](#).

### Qualitative Evaluation

A total of 95 comments were provided. Subsequently, wording of 7 items was refined in the checklist, and of 9 items, there were modifications in their corresponding Explanation & Elaboration section in [Multimedia Appendix 1](#). The evolutionary trajectory of the item list is presented in Figures S1 and S2 in [Multimedia Appendix 1](#).

### Final Reporting Item Checklist

[Table 2](#) presents the RATE-XR checklist and consists of 17 XR-specific reporting items (composed of 18 subitems) and 14 generic reporting items, selected by the Consensus Group.

**Table 2.** RATE-XR (reporting for the early-phase clinical evaluation of applications using extended reality) checklist.

Theme	Item number <sup>a</sup>	Recommendation
<b>Title and abstract</b>		
Title	1	<ul style="list-style-type: none"> <li>Identify the study as an early clinical evaluation, or a similar term, of an application using XR<sup>b</sup>, or a more specific term, in the title, including its intended aim.</li> </ul>
Abstract	I	<ul style="list-style-type: none"> <li>Provide a (structured) summary of the study.</li> <li>Consider including the following: <ul style="list-style-type: none"> <li>A concise description of the clinical problem or knowledge gap and the rationale for using an application using XR</li> <li>A concise description of the study methods, including a short description of the application including its name, study population, study setting, main outcomes, and assessment methods.</li> <li>A concise description of the results, including safety and harm outcomes</li> <li>A short conclusion</li> <li>If applicable, details about the registration of the study in a publicly available database.</li> </ul> </li> </ul>
<b>Introduction</b>		
Clinical problem and existing evidence	2	<ul style="list-style-type: none"> <li>Introduce the clinical problem for which the application using XR was used, including its relevance and a description of (the efficacy of) evidence-based or commonly used interventions or the treatment as usual, which is intended to be replaced by the application using XR.</li> </ul>
Introduction of the application	3	<ul style="list-style-type: none"> <li>Introduce the application using XR, including the following: <ul style="list-style-type: none"> <li>Hypotheses for the potential effect; how the application is expected to contribute to the clinical problem.</li> <li>If available, a concise description of, or a reference to, previous research on the same (or a similar) application.</li> </ul> </li> </ul>
Objectives	II	<ul style="list-style-type: none"> <li>Specify the study objectives or hypotheses.</li> </ul>
<b>Methods and analysis</b>		
Trial design and reporting	III	<ul style="list-style-type: none"> <li>Provide a reference to ethical approval and, if available, to any (published) study protocol and registration of the study in a publicly available repository.</li> </ul>
Trial design and reporting	IV	<ul style="list-style-type: none"> <li>Describe, and mention the rationale for, the study design. For clarification, it is recommended to use a flow diagram.</li> </ul>
Participants and setting	4	<ul style="list-style-type: none"> <li>Describe the setting and locations, including country, where data were collected and processed, and where the application using XR was applied and evaluated.</li> </ul>
Participants and setting	5a	<ul style="list-style-type: none"> <li>Describe how participants were selected and recruited and provide eligibility criteria.</li> </ul>
Participants and setting	5b	<ul style="list-style-type: none"> <li>Describe who will be applying the application and whether they were trained.</li> </ul>
Intervention and procedures	6	<ul style="list-style-type: none"> <li>Provide a description of the application, including its content, hardware, protocol, and set-up, or provide a reference to previous publications where this information is described. Consider supplementing the description with an image, figure, or film.</li> </ul>
Intervention and procedures	7	<ul style="list-style-type: none"> <li>Describe, or provide a reference to, the development process of the application.</li> </ul>
Intervention and procedures	8	<ul style="list-style-type: none"> <li>Describe the participant timeline in sufficient detail to allow replication, including all procedures, co-interventions (if applicable), and (follow-up) assessments.</li> </ul>
Intervention and procedures	V	<ul style="list-style-type: none"> <li>Describe and give a rationale for the control conditions or provide a rationale for not using one.</li> </ul>
Outcomes	VI	<ul style="list-style-type: none"> <li>Describe all prespecified primary and secondary outcomes, including how and when assessed.</li> </ul>
Outcomes	9	<ul style="list-style-type: none"> <li>Describe how safety and harm outcomes were assessed. Describe which, and how, other XR-specific outcomes were assessed, such as performance, usability, presence, perspectives, and acceptability.</li> </ul>
Sample size	VII	<ul style="list-style-type: none"> <li>Provide a justification for the sample size.</li> </ul>



Theme	Item number <sup>a</sup>	Recommendation
Analysis	VIII	<ul style="list-style-type: none"> <li>Provide a detailed description of how primary and secondary outcomes were analyzed, including any prespecified comparisons or stratifications.</li> </ul>
Protocol alterations	IX	<ul style="list-style-type: none"> <li>Describe changes to the methods or protocol, including procedures, study outcomes, eligibility criteria, and analysis plan, after study commencement, with reasons, and, if applicable, report whether the study registration was updated.</li> </ul>
<b>Results</b>		
Participant flow and recruitment	X	<ul style="list-style-type: none"> <li>Describe the time frame of recruitment and follow-up and the participant flow, including the number of patients screened and included, receiving the intervention, and being included in each analysis. Report if, and why, the study was prematurely terminated. The use of a flow diagram is highly recommended.</li> </ul>
Baseline data	XI	<ul style="list-style-type: none"> <li>Describe, or add a table depicting, baseline and treatment-related characteristics. If applicable, describe and specify any concurrent measures.</li> </ul>
Main results	XII	<ul style="list-style-type: none"> <li>Report on all prespecified outcomes that are available. Consider using tables, figures, or graphs to illustrate results.</li> </ul>
XR and human factors	10	<ul style="list-style-type: none"> <li>Include information about the usage of the application, such as duration, frequency, number of sessions, error rates, and number of sessions requiring interruption or discontinuation, including reasons.</li> </ul>
XR and human factors	11	<ul style="list-style-type: none"> <li>If assessed, report on XR-specific outcomes, such as performance, usability, presence, perspectives, and acceptability.</li> </ul>
Safety and harms	12	<ul style="list-style-type: none"> <li>Report on safety and harms, including unintended effects, both during and after using the application.</li> </ul>
<b>Discussion and conclusion</b>		
Generalizability and impact	13	<ul style="list-style-type: none"> <li>Discuss (potential) impact of study findings and generalizability, including barriers for the use and implementation of the application.</li> </ul>
Safety and harms	14	<ul style="list-style-type: none"> <li>Discuss safety and instances of harm, including their possible effects on study findings, implications for future use of the applications, and whether they can be prevented or mitigated.</li> </ul>
Ethics	15	<ul style="list-style-type: none"> <li>Describe ethical considerations, including benefits and risks, for the current and future use of the application.</li> </ul>
Strengths and limitations	XIII	<ul style="list-style-type: none"> <li>Discuss study strengths and limitations, including sources of potential bias.</li> </ul>
Conclusion	16	<ul style="list-style-type: none"> <li>Provide a conclusion that accurately interprets study findings, including future perspectives.</li> </ul>
<b>Statements</b>		
Funding and conflicts of interest	XIV	<ul style="list-style-type: none"> <li>Disclose any potential conflict of interest, real or apparent, including the funding sources and their roles in the design, conduct, analysis, and report of the study, potential roles of commercial companies, and personal conflicts of interest for each author.</li> </ul>
Application	17	<ul style="list-style-type: none"> <li>Indicate whether the application is a commercial product, it is publicly available, it can be accessed, it complies with the medical device regulations, and whether the application was approved for its intended use by a formal regulatory body or if the study is part of the clinical evaluation for future certification.</li> </ul>

<sup>a</sup>AI-specific items are numbered in Arabic numerals; generic items are numbered in Roman numerals.

<sup>b</sup>XR: extended reality.

## Discussion

### Reporting Item Checklist

The RATE-XR guideline serves as a checklist for reporting studies that focus on the early-phase evaluation of clinical applications using immersive technologies, regardless of the chosen study design (Figure 1). Depending on the specific study design selected, authors may also find it contributing to complement their reporting with guidelines tailored to that study type, such as the CONSORT guideline for randomized trials or the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guideline for observational studies [39,40]. This provides a helpful source to support researchers and reviewers assess manuscript compliance with the guideline. For a more elaborate understanding of each item's relevance and recommendations on how to report, we added an in-depth Explanation & Elaboration section for every item in Note 2 in Multimedia Appendix 1.

It is important to recognize that reporting guidelines, including the RATE-XR guideline, offer a framework of reporting recommendations, yet they may not comprehensively cover every aspect relevant to a particular study or guide the conduct of research. While not exhaustive, familiarity with RATE-XR can help researchers in the design and execution of studies within the guideline's scope. Given the challenge of reporting all required information into a single manuscript, authors may need to refer to other documents, such as study protocols, previous publications, and supplementary materials from digital repositories.

### Lessons Learned

The RATE-XR guideline represents the outcome of an extensive international consensus effort from a diverse and representative group of experts with a broad range of professional expertise and backgrounds. The high response rate and the remarkable level of engagement from stakeholders, along with the fact that 5 of the 7 most productive authors of XR research in psychology or medicine were willing to be included in the Steering Group, underscore the necessity for comprehensive reporting guidance in the early-phase clinical evaluation of XR applications [2]. This growing recognition highlights the increasing importance attributed to thorough clinical evaluation as a cornerstone for the effective implementation of XR technologies. The development of this reporting guideline was shaped by the Steering Group's belief that the use of XR-related health care applications will continue to expand, with an increasing requirement for high-quality, comprehensive, consistent, and generalizable reporting of early-phase evaluations.

The RATE-XR guideline is a pioneering endeavor, being the world's first reporting guideline specifically tailored to medical or psychology research involving XR-related applications. We focused on the early-phase clinical evaluation of XR-related applications, as existing guidelines insufficiently represented the essential reporting items for this type of research. Studies on late-phase evaluation often have the option to adhere to more general reporting guidelines, such as CONSORT for randomized controlled trials. However, we acknowledge that beyond this initial guideline, there is merit in further developing XR-specific

extensions for existing guidelines. Thus, our efforts mark the crucial first step in harmonizing XR research in the health care sector. Beyond its primary aim, the RATE-XR guideline may also serve as a compass for authors, guiding them in study design, protocol development, and the registration of early-phase studies involving XR applications.

To attract experts with diverse backgrounds across the health care field, we published a call for contributions in the form of a correspondence paper. In this publication, we mainly focused on the terminology "virtual reality," although we already had the more inclusive terminology "XR" in scope. During extensive deliberations during the Delphi rounds and after discussion within the Steering Group and Consensus Group meetings, we decided to use the more inclusive term XR instead of VR. This change allows the guideline to cover a broader range of applications, including augmented and MR, as well as any future immersive technologies. We concluded that XR terminology better represents all these applications, and all eventually will have to undertake equal and similar steps during early-stage evaluation and research. The decision to change the project's name from RATE-VR to RATE-XR reflects the guideline's adaptability and commitment to serving a diverse spectrum of applications beyond just focusing on VR alone.

The Delphi process, while invaluable for achieving consensus on guideline development, presented challenges such as maintaining a high follow-up rate among participants and reconciling diverse expert opinions. Our approach to addressing these challenges involved adaptive communication strategies and fostering an environment conducive to open dialogue, which were instrumental in enhancing participant engagement and consensus quality. These experiences provide key learnings that could inform similar guideline development efforts in emerging research fields.

Throughout the guideline's development, several topics generated more dynamic discussions than others, leading to a number of critical decisions. First, a discussion concerned the depth of information required about those administering XR applications. Some participants advocated for gathering baseline researcher or provider characteristics and offering detailed accounts of their training and qualifications. Given the novelty of the technology, the consensus was that mentioning that application providers had sufficient training was sufficient with no specific detailed requirement or baseline demographics unless deemed pivotal for study outcomes. In concordance, a recent review concluded that XR providers need training to improve adoption, which can be achieved using a variety of training programs, strategies, or educational resources, and that there is no minimum amount or golden standard of XR training [41].

Second, the deliberation on XR-specific outcomes, including factors such as acceptability, usability, user experiences, immersiveness, and cybersickness or other negative side-effects, sparked lengthy discussions among participants. While the importance of reporting cybersickness and safety-related aspects in early-phase evaluations was unanimously acknowledged for building trust within the research community, participants recognized the challenges of encompassing all XR-specific outcomes comprehensively in every article. Consequently, it

was agreed that cybersickness and safety should be mandatory reporting items, while other XR-specific outcomes should be considered optional, acknowledging the difficulty in fully addressing all these outcomes in every manuscript. This approach allows researchers the flexibility to focus on the most relevant outcomes for their specific studies.

Third, intensive debates centered around the level of detail necessary when describing XR application hardware, software, and development processes in the RATE-XR guideline. While some argued for comprehensive and mandatory disclosure, others championed flexibility. Ultimately, agreement was reached to consolidate these aspects into a single section in the guideline, with the Explanation & Elaboration note providing guidance on what to include and how to effectively report them.

Fourth, the need for items on randomization in the current guideline was discussed. Participants felt that most early-phase evaluations are seldom randomized and acknowledged that if a study has a randomized design, adhering to established guidelines such as CONSORT would be more appropriate than duplicating information in the RATE-XR guideline. It was agreed that the guideline should not delve into specific items related to randomization, as it would be more beneficial for researchers to consult CONSORT and ensure consistency in reporting across various study designs. A similar consideration and strategy was recently adopted in the DECIDE-AI guideline to prevent a too exhaustive reporting checklist in the early stage of innovative technology development [13].

Fifth, the topic of blinding and the utilization of control groups triggered significant debate during the consensus process. Participants recognized the inherent challenges of blinding in XR studies due to the immersive nature of the technology. While some argued that specific items on blinding-related items should be included, others emphasized the challenges in doing so effectively. Furthermore, in alignment with the guideline's aim to avoid duplicating items covered by study design-specific guidelines, the decision was made to omit the item on blinding. Nevertheless, control groups were deemed valuable for comparison purposes, resulting in the inclusion of items that address the presence or absence of control groups and outline

their characteristics as essential components within the guideline.

Sixth, discussions occurred regarding the necessity of justifying the sample sizes in the early-phase clinical assessment of XR applications. Participants held differing perspectives as to whether a formal sample size calculation should be mandated for all research types within the RATE-XR guideline. Ultimately, it was determined that while it is essential to provide some form of justification for the chosen sample size, not all research designs necessitate a formal sample size calculation. This decision recognizes the diversity of study designs and acknowledges that certain types of early-phase evaluations may have inherent limitations that preclude the use of traditional sample size calculations. Nevertheless, authors are strongly encouraged to provide a rationale and justification for their chosen sample size. This proactive step enhances transparency within the research process, enabling readers to assess the study's reliability of its findings.

Lastly, discussions concerning the inclusion of standardized statements such as data protection, ethics approval, and data and code availability were extensive. While their significance was acknowledged, the consensus was that most journals already require authors to address these elements in their manuscripts. Therefore, specific items related to these aspects were omitted from the RATE-XR guideline, as they are comprehensively covered by existing publication requirements and publication guidelines.

In conclusion, the RATE-XR guideline is a pioneering effort in facilitating comprehensive and standardized reporting in the early-phase clinical evaluation of XR applications. Its development involved extensive expert-informed debates and critical decisions that aim to ensure its purpose as a valuable resource for researchers while maintaining its adaptability to a dynamic and evolving field. We anticipate that this guideline will foster transparency, enhance the quality of reporting, and ultimately contribute to the responsible and effective integration of XR technologies in health care and related fields. Furthermore, we encourage the development of XR-specific extensions for existing guidelines to further advance the harmonization of XR research practices.

## Acknowledgments

There was no funding in any form for this study. We have not received payments from any agency or pharmaceutical company to establish this manuscript.

The members of RATE-XR Expert Group are Aisling Flynn, Alice Chirico, Amir H Sadeghi, Andrea Gaggioli, Andrea S Won, Annelotte P van Haaps, Azucena Garcia-Palacios, Beate Dejacó, Bram Dierckx, Carsten Finke, Catheline van Driel, Chris NW Geraets, Christopher Eccleston, Christopher R Madan, Clarine van Oel, Constantinos Panayi, Darcy Ummels, Dave Thomas, Debra L Safer, Elsbeth Zandee, Emil R Høeg, Felix J Hüttner, Floris van der Breggen, Geert-Jan van Geffen, Guido A Hakvoort, Giulia Corno, Hadi Hosseini, Hafize Demirci, Hanne Konradsen, Henry Xiang, Ivan Phelan, Jan D Rölfling, Jan Gödeke, Jennifer N Stinson, Jeremy Bailenson, Jeroen Legerstee, Jiabin Shen, Joost Huiskens, Jordan Tsigarides, Jose F Costa, Juana M Bretón-López, Karamveer Narang, Karin Valkenet, Kim D Bullock, Kimmy Rosielle, Krista Hoek, Line K Pedersen, Loes Bulle-Smid, Lonneke M Staals, Kemal Kuscu, Mareine GT Koornneef, Margaux Sageot, Margot D Paul, Maria Bajwa, Maria Matsangidou, Marie-Madlen Jeitziner, Mariju F Baluyot, Marlies P Schijven, Martine J van Bennekom, Matthew Browning, Melissa L Morris, Merel A Oskam, Merlijn Smits, Michael Gaebler, Mienke Rijdsdijk, Njin-Zu Chen, Omar Aly, Pablo Campo-Prieto, Panagiotis Kourtesis, Philipp Kellmeyer, Les Posen, Rachel Reeves, Rami A Ahmed, Raphael R Bruno, Rob JEM Smeets, Robbert Brouwer, Robert J Fine, Robert M Lundin, Roderick F van Beek, Rosa M Baños, Roselinde MCA Pot-Kolder,

Sarah E MacPherson, Silvia Serino, Sophia Rekers, Srinivasan S Pillay, Stephan Krohn, Stéphane Bouchard, Sulayman el Mathari, Susan Persky, Syl Slatman, Synthia Guimond, Thomas J Caruso, Thomas Sauter, Thomas Wolbers, Tjitske D Groenveld, Tobias Loetscher, Todd Chang, Tonnie Staring, Vishnunarayan G Prabhu, Wim Veling, and Winnie WS Mak.

### Data Availability

The datasets and necessary documents generated and analyzed during this study are available in [Multimedia Appendix 1](#) and the digital Open Science Framework RATE-XR platform [34]. Additionally, all study materials are available from the corresponding author upon reasonable request.

### Authors' Contributions

MEvG initiated the study. JHV, JvB, and MEvG designed the study. Members of the RATE-XR Steering Group (DD, JvB, GR, ASR, LH, BG, EJW, and DG) provided methodological input and oversaw the conduct of the study. JHV and DD conducted the thematic analysis and Delphi rounds analysis and produced the Delphi round summaries. All members of the steering group (JHV, DD, JvB, GR, BKW, PC, ASR, BOR, CB, LH, OJB, CJ, BG, EJW, BJB, DG, and MEvG) selected the final content and wording of the guidelines. JHV, DD, and MEvG chaired the consensus meeting. JHV, DD, and MEvG drafted the final manuscript and Explanation & Elaboration note. All authors reviewed and commented on the final manuscript and Explanation & Elaboration note. All members of the steering group collaborated in the development of the guidelines by participating in the Delphi process, the qualitative evaluation of the guidelines, or both.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

The RATE-XR (reporting for the early-phase clinical evaluation of applications using extended reality) qualitative study guideline. [[DOCX File , 335 KB-Multimedia Appendix 1](#)]

### References

1. Topol E. The Topol Review: Preparing the Healthcare Workforce to Deliver the Digital Future. England. Health Education England; 2019:104. URL: <https://topol.hee.nhs.uk/>
2. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. *J Med Internet Res*. 2021;23(2):e25499. [[FREE Full text](#)] [doi: [10.2196/25499](https://doi.org/10.2196/25499)] [Medline: [33565986](https://pubmed.ncbi.nlm.nih.gov/33565986/)]
3. Bruno RR, Bruining N, Jung C, VR-ICU Study group. Virtual reality in intensive care. *Intensive Care Med*. 2022;48(9):1227-1229. [[FREE Full text](#)] [doi: [10.1007/s00134-022-06792-0](https://doi.org/10.1007/s00134-022-06792-0)] [Medline: [35816236](https://pubmed.ncbi.nlm.nih.gov/35816236/)]
4. Kanschik D, Bruno RR, Wolff G, Kelm M, Jung C. Virtual and augmented reality in intensive care medicine: a systematic review. *Ann Intensive Care*. 2023;13(1):81. [[FREE Full text](#)] [doi: [10.1186/s13613-023-01176-z](https://doi.org/10.1186/s13613-023-01176-z)] [Medline: [37695464](https://pubmed.ncbi.nlm.nih.gov/37695464/)]
5. Cipresso P, Giglioli IAC, Raya MA, Riva G. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Front Psychol*. 2018;9:2086. [[FREE Full text](#)] [doi: [10.3389/fpsyg.2018.02086](https://doi.org/10.3389/fpsyg.2018.02086)] [Medline: [30459681](https://pubmed.ncbi.nlm.nih.gov/30459681/)]
6. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. *Cochrane Database Syst Rev*. 2017;11(11):CD008349. [[FREE Full text](#)] [doi: [10.1002/14651858.CD008349.pub4](https://doi.org/10.1002/14651858.CD008349.pub4)] [Medline: [29156493](https://pubmed.ncbi.nlm.nih.gov/29156493/)]
7. Jung C, Wolff G, Wernly B, Bruno RR, Franz M, Schulze PC, et al. Virtual and augmented reality in cardiovascular care: state-of-the-art and future perspectives. *JACC Cardiovasc Imaging*. 2022;15(3):519-532. [[FREE Full text](#)] [doi: [10.1016/j.jcmg.2021.08.017](https://doi.org/10.1016/j.jcmg.2021.08.017)] [Medline: [34656478](https://pubmed.ncbi.nlm.nih.gov/34656478/)]
8. Bruno RR, Wolff G, Wernly B, Masyuk M, Piayda K, Leaver S, et al. Virtual and augmented reality in critical care medicine: the patient's, clinician's, and researcher's perspective. *Crit Care*. 2022;26(1):326. [[FREE Full text](#)] [doi: [10.1186/s13054-022-04202-x](https://doi.org/10.1186/s13054-022-04202-x)] [Medline: [36284350](https://pubmed.ncbi.nlm.nih.gov/36284350/)]
9. Tsai TY, Onuma Y, Złahoda-Huzior A, Kageyama S, Dudek D, Wang Q, et al. Merging virtual and physical experiences: extended realities in cardiovascular medicine. *Eur Heart J*. 2023;44(35):3311-3322. [[FREE Full text](#)] [doi: [10.1093/eurheartj/ehad352](https://doi.org/10.1093/eurheartj/ehad352)] [Medline: [37350487](https://pubmed.ncbi.nlm.nih.gov/37350487/)]
10. Rizzo AS, Koenig ST. Is clinical virtual reality ready for primetime? *Neuropsychology*. 2017;31(8):877-899. [doi: [10.1037/neu0000405](https://doi.org/10.1037/neu0000405)] [Medline: [29376669](https://pubmed.ncbi.nlm.nih.gov/29376669/)]
11. Hirst A, Philippou Y, Blazeby J, Campbell B, Campbell M, Feinberg J, et al. No surgical innovation without evaluation: evolution and further development of the IDEAL framework and recommendations. *Ann Surg*. 2019;269(2):211-220. [doi: [10.1097/SLA.0000000000002794](https://doi.org/10.1097/SLA.0000000000002794)] [Medline: [29697448](https://pubmed.ncbi.nlm.nih.gov/29697448/)]

12. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, Balliol Collaboration, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374(9695):1105-1112. [doi: [10.1016/S0140-6736\(09\)61116-8](https://doi.org/10.1016/S0140-6736(09)61116-8)] [Medline: [19782876](https://pubmed.ncbi.nlm.nih.gov/19782876/)]
13. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. DECIDE-AI expert group. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. 2022;28(5):924-933. [doi: [10.1038/s41591-022-01772-9](https://doi.org/10.1038/s41591-022-01772-9)] [Medline: [35585198](https://pubmed.ncbi.nlm.nih.gov/35585198/)]
14. Beams R, Brown E, Cheng W, Joyner JS, Kim AS, Kontson K, et al. Evaluation challenges for the application of extended reality devices in medicine. *J Digit Imaging*. 2022;35(5):1409-1418. [FREE Full text] [doi: [10.1007/s10278-022-00622-x](https://doi.org/10.1007/s10278-022-00622-x)] [Medline: [35469355](https://pubmed.ncbi.nlm.nih.gov/35469355/)]
15. Fleetcroft C, McCulloch P, Campbell B. IDEAL as a guide to designing clinical device studies consistent with the new European medical device regulation. *BMJ Surg Interv Health Technol*. 2021;3(1):e000066. [FREE Full text] [doi: [10.1136/bmjst-2020-000066](https://doi.org/10.1136/bmjst-2020-000066)] [Medline: [35047800](https://pubmed.ncbi.nlm.nih.gov/35047800/)]
16. Birkhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. *JMIR Ment Health*. 2019;6(1):e11973. [FREE Full text] [doi: [10.2196/11973](https://doi.org/10.2196/11973)] [Medline: [30702436](https://pubmed.ncbi.nlm.nih.gov/30702436/)]
17. Kellmeyer P, Biller-Andorno N, Meynen G. Ethical tensions of virtual reality treatment in vulnerable patients. *Nat Med*. 2019;25(8):1185-1188. [doi: [10.1038/s41591-019-0543-y](https://doi.org/10.1038/s41591-019-0543-y)] [Medline: [31359003](https://pubmed.ncbi.nlm.nih.gov/31359003/)]
18. Omoumi P, Ducarouge A, Tournier A, Harvey H, Kahn CE, Louvet-de Verchère F, et al. To buy or not to buy-evaluating commercial AI solutions in radiology (the ECLAIR guidelines). *Eur Radiol*. 2021;31(6):3786-3796. [FREE Full text] [doi: [10.1007/s00330-020-07684-x](https://doi.org/10.1007/s00330-020-07684-x)] [Medline: [33666696](https://pubmed.ncbi.nlm.nih.gov/33666696/)]
19. Lewis CH, Griffin MJ. Human factors consideration in clinical applications of virtual reality. *Stud Health Technol Inform*. 1997;44:35-56. [Medline: [10175342](https://pubmed.ncbi.nlm.nih.gov/10175342/)]
20. Riva G, Wiederhold BK. What the metaverse is (really) and why we need to know about it. *Cyberpsychol Behav Soc Netw*. 2022;25(6):355-359. [doi: [10.1089/cyber.2022.0124](https://doi.org/10.1089/cyber.2022.0124)] [Medline: [35696299](https://pubmed.ncbi.nlm.nih.gov/35696299/)]
21. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217. [FREE Full text] [doi: [10.1371/journal.pmed.1000217](https://doi.org/10.1371/journal.pmed.1000217)] [Medline: [20169112](https://pubmed.ncbi.nlm.nih.gov/20169112/)]
22. Humphrey-Murto S, Wood TJ, Gonsalves C, Mascioli K, Varpio L. The Delphi method. *Acad Med*. Jan 2020;95(1):168. [doi: [10.1097/ACM.0000000000002887](https://doi.org/10.1097/ACM.0000000000002887)] [Medline: [31335812](https://pubmed.ncbi.nlm.nih.gov/31335812/)]
23. Dalkey N, Helmer O. An experimental application of the DELPHI method to the use of experts. *Manage Sci*. 1963;9(3):458-467. [doi: [10.1287/mnsc.9.3.458](https://doi.org/10.1287/mnsc.9.3.458)]
24. Vlake JH, Wils E, van Bommel J, Korevaar TIM, Gommers D, van Genderen ME. Virtual reality tailored to the needs of post-ICU patients: a safety and immersiveness study in healthy volunteers. *Crit Care Explor*. 2021;3(5):e0388. [FREE Full text] [doi: [10.1097/CCE.0000000000000388](https://doi.org/10.1097/CCE.0000000000000388)] [Medline: [34079940](https://pubmed.ncbi.nlm.nih.gov/34079940/)]
25. Vlake JH, Van Bommel J, Wils E, Korevaar TIM, Bienvenu OJ, Klijn E, et al. Virtual reality to improve sequelae of the postintensive care syndrome: a multicenter, randomized controlled feasibility study. *Crit Care Explor*. 2021;3(9):e0538. [FREE Full text] [doi: [10.1097/CCE.0000000000000538](https://doi.org/10.1097/CCE.0000000000000538)] [Medline: [34549192](https://pubmed.ncbi.nlm.nih.gov/34549192/)]
26. Vlake JH, Wils EJ, van Bommel J, Gommers D, van Genderen ME, HORIZON-ICU study group. Familiarity with the post-intensive care syndrome among general practitioners and opportunities to improve their involvement in ICU follow-up care. *Intensive Care Med*. 2022;48(8):1090-1092. [doi: [10.1007/s00134-022-06782-2](https://doi.org/10.1007/s00134-022-06782-2)] [Medline: [35776162](https://pubmed.ncbi.nlm.nih.gov/35776162/)]
27. Vlake JH, van Bommel J, Wils E, Bienvenu J, Hellemons ME, Korevaar TI, et al. Intensive care unit-specific virtual reality for critically ill patients with COVID-19: multicenter randomized controlled trial. *J Med Internet Res*. 2022;24(1):e32368. [FREE Full text] [doi: [10.2196/32368](https://doi.org/10.2196/32368)] [Medline: [34978530](https://pubmed.ncbi.nlm.nih.gov/34978530/)]
28. Bilbro NA, Hirst A, Paez A, Vasey B, Pufulete M, Sedrakyan A, et al. IDEAL Collaboration Reporting Guidelines Working Group. The IDEAL reporting guidelines: a Delphi consensus statement stage specific recommendations for reporting the evaluation of surgical innovation. *Ann Surg*. 2021;273(1):82-85. [doi: [10.1097/SLA.0000000000004180](https://doi.org/10.1097/SLA.0000000000004180)] [Medline: [32649459](https://pubmed.ncbi.nlm.nih.gov/32649459/)]
29. Dockx K, Bekkers EM, Van den Bergh V, Ginis P, Rochester L, Hausdorff JM, et al. Virtual reality for rehabilitation in Parkinson's disease. *Cochrane Database Syst Rev*. 2016;12(12):CD010760. [FREE Full text] [doi: [10.1002/14651858.CD010760.pub2](https://doi.org/10.1002/14651858.CD010760.pub2)] [Medline: [28000926](https://pubmed.ncbi.nlm.nih.gov/28000926/)]
30. Clinical Evaluation. IMDRF Medical Device Clinical Evaluation Form. 2019. URL: <https://tinyurl.com/5n6jxst9> [accessed 2024-09-27]
31. Executive Summary for the Patient Engagement Advisory Meeting: Augmented Reality and Virtual Reality Medical Device. 2022. URL: <https://www.fda.gov/media/159709/download>: [accessed 2024-09-27]
32. Public Workshop - Medical Extended Reality: Towards Best Evaluation Practices for Virtual And Augmented Reality in Medicine. 2020. URL: <https://www.fda.gov/media/159709/download>: [accessed 2024-09-27]
33. Vlake JH, van Bommel J, Riva G, Wiederhold BK, Cipresso P, Rizzo AS, et al. Reporting the early stage clinical evaluation of virtual-reality-based intervention trials: RATE-VR. *Nat Med*. 2023;29(1):12-13. [doi: [10.1038/s41591-022-02085-7](https://doi.org/10.1038/s41591-022-02085-7)] [Medline: [36526724](https://pubmed.ncbi.nlm.nih.gov/36526724/)]

34. Vlake JH, Drop DLQ, van Genderen ME. Open Science Framework RATE-XR. 2023. URL: <https://osf.io/cw6yq/> [accessed 2024-09-27]
35. Warström J. Mentimeter. URL: <https://www.mentimeter.com/> [accessed 2023-06-15]
36. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health*. 2020;2(10):e537-e548. [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30218-1](https://doi.org/10.1016/S2589-7500(20)30218-1)] [Medline: [33328048](https://pubmed.ncbi.nlm.nih.gov/33328048/)]
37. International Organization for Standardization. Information technology - artificial intelligence (AI) - bias in AI systems and AI aided decision making. URL: <https://www.iso.org/standard/77607.html> [accessed 2024-09-27]
38. International Organization for Standardization. Ergonomics of human-system interaction—part 11: usability: definitions and concepts. 2018. URL: <https://www.iso.org/> [accessed 2024-09-27]
39. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Br Med J*. 2010;340:c332. [FREE Full text] [doi: [10.1136/bmj.c332](https://doi.org/10.1136/bmj.c332)] [Medline: [20332509](https://pubmed.ncbi.nlm.nih.gov/20332509/)]
40. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. STROBE Initiative. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Br Med J*. 2007;335(7624):806-808. [doi: [10.1136/bmj.39335.541782.AD](https://doi.org/10.1136/bmj.39335.541782.AD)] [Medline: [17947786](https://pubmed.ncbi.nlm.nih.gov/17947786/)]
41. 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;4(1):67. [doi: [10.1186/s43058-023-00442-2](https://doi.org/10.1186/s43058-023-00442-2)] [Medline: [37328858](https://pubmed.ncbi.nlm.nih.gov/37328858/)]

## Abbreviations

**AI:** artificial intelligence

**AR:** augmented reality

**CONSORT:** Consolidated Standards of Reporting Trials

**EQUATOR:** Enhancing the Quality and Transparency of Health Research

**MR:** mixed reality

**OSF:** Open Science Framework

**RATE-XR:** reporting for the early-phase clinical evaluation of applications using extended reality

**SPIRIT:** Standard Protocol Items: Recommendations for Interventional Trials

**STROBE:** Strengthening the Reporting of Observational Studies in Epidemiology

**VR:** virtual reality

**XR:** extended reality

*Edited by N Cahill, T Leung; submitted 26.01.24; peer-reviewed by J Peek, X Liu; comments to author 25.04.24; revised version received 03.06.24; accepted 11.09.24; published 29.11.24*

*Please cite as:*

*Vlake JH, Drop DLQ, Van Bommel J, Riva G, Wiederhold BK, Cipresso P, Rizzo AS, Rothbaum BO, Botella C, Hooft L, Bienvenu OJ, Jung C, Geerts B, Wils E-J, Gommers D, van Genderen ME, RATE-XR Expert Group*

*Reporting Guidelines for the Early-Phase Clinical Evaluation of Applications Using Extended Reality: RATE-XR Qualitative Study Guideline*

*J Med Internet Res* 2024;26:e56790

URL: <https://www.jmir.org/2024/1/e56790>

doi: [10.2196/56790](https://doi.org/10.2196/56790)

PMID:

©Johan H Vlake, Denzel LQ Drop, Jasper Van Bommel, Giuseppe Riva, Brenda K Wiederhold, Pietro Cipresso, Albert S Rizzo, Barbara O Rothbaum, Cristina Botella, Lotty Hooft, Oscar J Bienvenu, Christian Jung, Bart Geerts, Evert-Jan Wils, Diederik Gommers, Michel E van Genderen, RATE-XR Expert Group. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 29.11.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the *Journal of Medical Internet Research* (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.