Tutorial

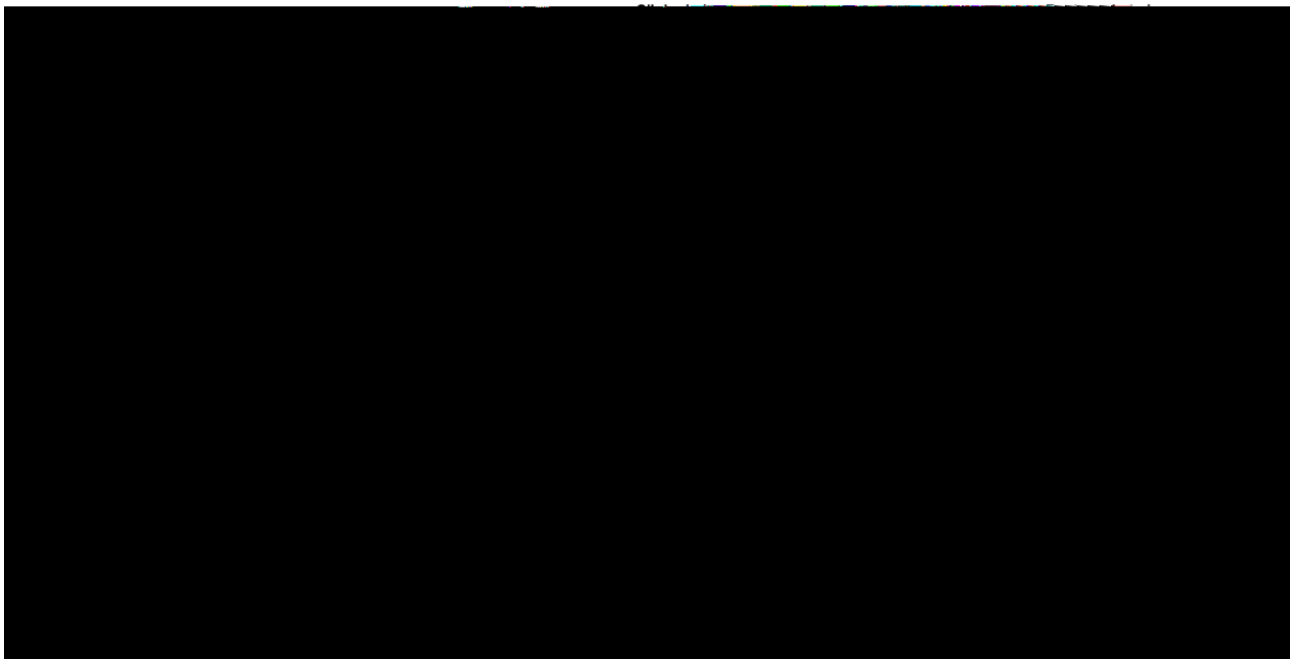
Reporting Guidelines for the Early-Phase Clinical Evaluation of Applications Using Extended Reality: RATE-XR Qualitative Study Guideline

Johan H Vlake[1,2], BSc; Denzel L Q Drop[1], MD; Jasper Van Bommel[1], MD, PhD; Giuseppe Riva[3,4], PhD; Brenda K Wiederhold[5], BCN, BCB, MBA, PhD; Pietro Cipresso[3,6], PhD; Albert S Rizzo[7], PhD; Barbara O Rothbaum[8], PhD; Cristina Botella[9,10], PhD; Lotty Hooft[11], PhD; Oscar J Bienvenu[12], MD, PhD; Christian Jung[13,14], MD,

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

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

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

Table 1.

(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.

XSL•FO
RenderX

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

Theme	Item number[a]	Recommendation
Title and abstract		
Title	1	• Identify the study as an early clinical evaluation, or a similar term, of an application using XR[b], or a more specific term, in the title, including its intended aim.

Abstract	I	• Provide a (structured) summary of the study.
		• Consider including the following:
		  • A concise description of the clinical problem or knowledge gap and the rationale for using an application using XR
		  • 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.
		  • A concise description of the results, including safety and harm outcomes
		  • A short conclusion
		  • If applicable, details about the registration of the study in a publicly available database.

Introduction		
Clinical problem and existing evidence	2	• 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.

Introduction of the application	3	• Introduce the application using XR, including the following:
		  • Hypotheses for the potential effect; how the application is expected to contribute to the clinical problem.
		  • If available, a concise description of, or a reference to, previous research on the same (or a similar) application.

Objectiv	II	

Theme	Item number[a]	Recommendation
Analysis	VIII	• Provide a detailed description of how primary and secondary outcomes were analyzed, including any prespecified comparisons or stratifications.
Protocol alterations	IX	• 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.
Results		
Participant flow and recruitment	X	• 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.
Baseline data	XI	• Describe, or add a table depicting, baseline and treatment-related characteristics. If applicable, describe and specify any concurrent measures.
Main results	XII	• Report on all prespecified outcomes that are available. Consider using tables, figures, or graphs to illustrate results.
XR and human factors	10	• 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.
XR and human factors	11	• If assessed, report on XR-specific outcomes, such as performance, usability, presence, perspectives, and acceptability.
		• Report on safety XR Tm(ic)Tj1 05 312.stion,

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

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

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.

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] [Medline: 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] [Medline: 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] [Medline: 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/bmjsit-2020-000066] [Medline: 35047800]
16. Birckhead 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] [Medline: 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] [Medline: 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] [Medline: 33666696]
19. Lewis CH, Griffin MJ. Human factors consideration in clinical applications of virtual reality. Stud Health Technol Inform.

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] [Medline: 33328048]
37. International Organization for Standardization. Information technology - artificial