Self-guided versus facilitator-guided debriefing in immersive virtual reality simulation: Protocol for a randomized controlled non-inferiority trial assessing teamwork skills in medical students

Amalie Middelboe Sohlin[1]*, Anja Poulsen[1], Ida Madeline Hoffmann[1], Line Klingen Gjærde[1,2], Stine Lund[3,4], Gritt Overbeck[5], Lone Paulsen[6], Todd P. Chang[7], Joy Yeonjoo Lee[8], Jette Led Sørensen[2,4], Jesper Kjærgaard[1]

1 Department of Paediatrics and Adolescent Medicine, Copenhagen University Hospital – Rigshospitalet, Copenhagen, Denmark, 2 Mary Elizabeth's Hospital and Juliane Marie Centre, Copenhagen University Hospital – Rigshospitalet, Copenhagen, Denmark, 3 Department of Pediatrics, Copenhagen University Hospital North Zealand, Hillerød, Denmark, 4 Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, 5 Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, 6 H.C. Andersen Children's Hospital, Odense University Hospital, Odense, Denmark, 7 Children's Hospital Los Angeles, Keck School of Medicine of University of Southern California, Los Angeles, California, United States of America, 8 Faculty of Governance and Global Affairs, The Hague, Leiden University, Netherlands

* amalie.middelboe.andersen@regionh.dk

Abstract

Simulation-based medical education has been shown to be more effective but also logistically demanding and costly compared to other educational strategies in developing medical skills. Immersive virtual reality is an emerging technology enabling learners to train without a facilitator through computer-generated feedback, offering the potential for increased flexibility in the timing and location of the training and reduced costs. However, little is known about whether immersive virtual reality simulation yields similar results with and without a facilitator. The aim of this study is to compare the effects of self-guided compared to facilitator-guided debriefing for immersive virtual reality simulation-based pediatric emergency team training. We will conduct a randomized, controlled, single-blinded non-inferiority study with a parallel group, pretest-post-test design. 88 medical students (44 teams) will be randomized to undergo immersive virtual reality simulation-based pediatric emergency team training with either self-guided or facilitator-guided debriefing. We will assess the teams before and after the virtual reality intervention in a mannequin-based simulation. The mannequin-based simulation will be videorecorded, and two independent raters, blinded to group allocation, will assess the recordings using validated scales measuring teamwork skills (primary outcome), ABCDE adherence, and time to critical actions. We will further collect data on perceptions of debriefing quality, motivation, workload, usability, and cybersickness. To account for repeated measures and clustering within teams, we will apply a linear mixed model for data analysis. This study

aims to provide insight into the effects of self-guided versus facilitator-guided debriefing in immersive virtual reality simulation, with implications for the future development and implementation of immersive virtual reality simulation in medical education.

We have registered the trial on ClinicalTrials.gov (identifier: NCT06956833).

Introduction

Simulation-based medical education has been shown to significantly enhance the skills of healthcare professionals and students, as well as improve patient outcomes [1–6]. Several systematic reviews and meta-analyses have demonstrated that simulation-based medical education outperforms traditional clinical education and other educational strategies in improving a wide range of medical skills [4,5]. However, simulation-based medical education is also more logistically demanding and costly compared to many other educational strategies [5], prompting calls for research into strategies that maximize its benefits while increasing access and minimizing costs [7].

Immersive virtual reality (VR) is an emerging technology within simulation-based medical education [8–11]. Recent research suggests that VR can provide learning outcomes comparable to other simulation modalities, while potentially offering greater flexibility in training schedules and reducing costs [9,11]. One of its main advantages in enhancing flexible use and reducing costs is the potential reduction in educational professionals needed to facilitate the simulations, as VR scenarios can be programmed to adapt dynamically to learners' actions, allowing learners to train independently of a facilitator [12]. However, little is known about whether the presence of a facilitator significantly influences the effectiveness of VR simulation.

In simulation-based medical education, the facilitator plays a key role in guiding the debriefing conversation [13]. Debriefing, described as a structured reflection aiming to explore and learn from the simulation-based experience [14,15], has been shown to significantly enhance clinical performance outcomes [16–20]. While facilitator-guided debriefing is widely regarded as the gold standard [13], learners can also lead their own reflection through *self-guided debriefing* [14]. In such cases, the debriefing is typically supported by cognitive aids, such as recordings of the simulation or a debriefing framework, to support self- and peer-assessment [14,21].

The self-guided approaches are grounded in theories of self-regulated learning, such as self-determination theory, which suggest that autonomy may enhance learner engagement and motivation [22]. Research suggests that self-guided approaches may offer additional benefits, including increased ownership and commitment to change [23], flexibility in the duration of the training [24,25], and greater motivation for self-regulated learning [26]. However, learners with lower self-regulation skills may struggle to evaluate their own performance, perceive the process as less effective, be less motivated, and experience increased cognitive load [25,27,28].

While self-guided and facilitator-guided debriefing have been compared in other domains of simulation-based medical education, such as mannequin- and

screen-based simulation [18,21,29], little is known about the effectiveness of self-guided debriefing in VR [30]. For VR to fulfil its potential as an independent and scalable simulation-based medical education modality, it is essential to determine whether self-guided debriefing can match the effectiveness facilitator-guided debriefing in a VR context.

Previous studies on immersive VR in simulation-based emergency training

Little is known about the effect of immersive VR on teamwork skills. This represents a significant gap in the literature, as effective teamwork and communication are critical to preventing patient harm [31–33]. In the field of emergency training, VR has shown promising results for improving the individual skills of healthcare professionals and students, such as clinical reasoning and task prioritization [34–38], triage in mass causality incidents [39], memorization of the ABCDE approach [40], and neonatal resuscitation skills [41]. However, there has been a recent call for research on the effect of VR on *team-based* skill performance and process measures [10]. Further, while one of the proposed key advantages of VR is the possibility of facilitator-independent training, little is known about the role of the facilitator and different debriefing methods in VR for emergency training [10].

Pilot study

We conducted a randomized, controlled, single-blinded pilot study with parallel group, pretest-posttest design from September to December 2023 [30]. We randomized 24 healthcare professionals (trainee doctors and nurses) from four pediatric units at Copenhagen University Hospital, Rigshospitalet, to participate in VR simulation-based pediatric emergency team training with either self-guided or facilitator-guided debriefing. Teams were assessed at baseline and at one month follow-up in a mannequin-based pediatric emergency simulation. Two independent raters blinded to group allocation assessed team performance based on video-recordings of the mannequin-based simulations. Participants completed questionnaires on their perceptions of debriefing quality, motivation, usability, and workload.

Both interventions were found acceptable and usable. However, nearly half of the participants reported some degree of cybersickness, prompting recommendations to reduce VR exposure, ensure breaks between immersions, and optimize headset alignment. Additionally, coordinating clinically working healthcare professionals across four pediatric units to participate at across three time points proved logistically challenging. To support recruitment feasibility and reduce loss to follow-up in a definitive trial, we decided to limit data collection time points and include participants with more flexible schedules, such as students.

Preliminary results showed significant improvements from baseline to follow-up in both groups for teamwork skills (facilitator-guided mean 1.2, confidence interval (CI) 0.5 to 1.9, p = 0.005; self-guided mean 1.4, CI 0.8 to 2.0, p < 0.001) and ABCDE adherence (facilitator-guided mean 0.5, CI 0.1 to 1.0, p = 0.02; self-guided mean 0.5, CI 0.1 to 0.9, p = 0.01). The facilitator-guided group rated the debriefing quality higher (mean difference 1.4, CI 0.1 to 2.7, p = 0.04), but usability, workload, and motivation were similar across debriefing groups. We concluded that a study designed for testing a non-inferiority hypothesis can be relevant for conclusive evidence.

Aims and research questions

The aim of this study is to evaluate the effects of immersive VR simulation-based pediatric emergency team training with self-guided versus facilitator-guided debriefing in a randomized, controlled, non-inferiority study with parallel-group, pretest-posttest design.

For the primary outcome, we will investigate the following research question:

Research question 1 (RQ1 – effect on teamwork skills)

Does VR-based pediatric emergency team training with self-guided debriefing yield a non-inferior effect on medical students' non-technical team performance (teamwork skills) compared to VR with facilitator-guided debriefing?

For the secondary outcomes, we will investigate the following research questions:

Research question 2 (RQ2 – effect on overall team performance)

Do the two interventions result in a similar effect on medical students' technical team performance (ABCDE adherence and time to critical actions)? (RQ2a)

Does VR improve overall team performance (teamwork skills, ABCDE adherence, and time to critical actions) in both intervention groups? (RQ2b)

Research question 3 (RQ3 – perceived effectiveness and engagement)

Do the two interventions result in a similar engagement in VR simulation and debriefing (intrinsic motivation and perceived debriefing quality)?

Research question 4 (RQ4 – cognitive load)

Do the two interventions result in a similar cognitive load?

Research question 5 (RQ5 – usability and cybersickness)

What is the incidence of cybersickness and the perceived usability of VR in the two interventions?

Methods

Setting and population

The study will be conducted at Copenhagen University Hospital, Rigshospitalet, and Odense University Hospital, Denmark. Eligible participants will be medical students enrolled at Faculty of Health and Medical Sciences, Copenhagen University and University of Southern Denmark, who are within two years of graduation. Exclusion criteria will be lack of informed consent.

Participants will be recruited between May and October 2025 via emails sent to the consultants responsible for medical students' clinical placements at pediatric departments in the Capital Region of Denmark, Region Zealand, and the Region of Southern Denmark. These consultants will be encouraged to inform medical students, both in person and via email, about the opportunity to participate in the study. Additionally, information and invitations to participate will be shared on social media platforms targeting medical students in the relevant regions. Written informed consent, including permission to record and store all video data for the research project, will be obtained from all participants upon enrolment.

Study design

We will conduct a single-blinded, randomized controlled non-inferiority study with parallel-group, pretest-posttest design, following the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statement on clinical trial protocols [42].

Teams of medical students will be randomized at a 1:1 allocation ratio to go through immersive VR simulation-based pediatric emergency team training with self-guided or facilitator-guided debriefing. To evaluate the effect of the interventions on team performance, the teams will be assessed managing a mannequin-based pediatric emergency before and after the intervention (Fig 1). The mannequin-based simulation will be video recorded, and two independent observers blinded to group allocation will rate the collective performance of each team based on the videos.

Interventions

Both interventions will last four hours, including one hour of immersive VR simulation and one hour of debriefing (Fig 2). To minimize cybersickness, participants will spend at least 30 minutes outside the VR environment between scenarios. If cybersickness symptoms occur, they will be prompted to exit the simulation and continue by observing on a screen, providing supervision to their teammate.

In the facilitator-guided group, a medical doctor with formal training and experience in simulation and debriefing will guide the debriefings using the same VR-modified PEARLS framework as the self-guided group. This ensures that the only planned difference between groups is the presence or absence of a facilitator [30,43].

	STUDY PERIOD				
	Enrolment	Allocation	Post-randomization		
TIMEPOINT*	-t1 to 0	0	t1	t2	t3
ENROLMENT:					
Eligibility screen	X				
Informed consent	X				
Randomization		X			
INTERVENTIONS:					
VR with self-guided debriefing				X	
VR with facilitator-guided debriefing				X	
ASSESSMENTS:					
Sociodemographic data	X		X		
CTS			X		X
ABCDE			X		X
Time to critical action			X		X
DASH-SV				X	
IMI					X
NASA-TLX					X
SUS					X
VRSQ					X

Fig 1. Participant timeline: Schedule of enrolment, interventions, and assessments as recommended by the 2025 SPIRIT statement. * Enrolment and allocation will take place 1–8 weeks prior to t1. t1, t2, and t3 will be conducted in the morning, midday, and afternoon of the same day, respectively. ABCDE: airways, breathing, circulation, disability, and exposure; CTS: clinical teamwork scale; DASH-SV: debriefing assessment for simulation in healthcare student version; IMI: intrinsic motivation inventory; NASA-TLX: national aeronautics and space administration task load index; SUS: system usability scale; VR: virtual reality; VRSQ: virtual reality sickness questionnaire.

https://doi.org/10.1371/journal.pone.0332309.g001

Development of immersive VR pediatric emergency scenarios and core game mechanics

Three immersive VR scenarios were developed and tested in our pilot study [30]. The scenarios depict three pediatric emergency cases: sepsis, respiratory distress, and anaphylaxis. Interactive elements are embedded in the virtual environment and trigger predefined changes in the patient's vitals and clinical condition. Dropdown menus are used to simulate palpating of the patient. Participants will navigate the emergency room via teleportation. The cases were developed by
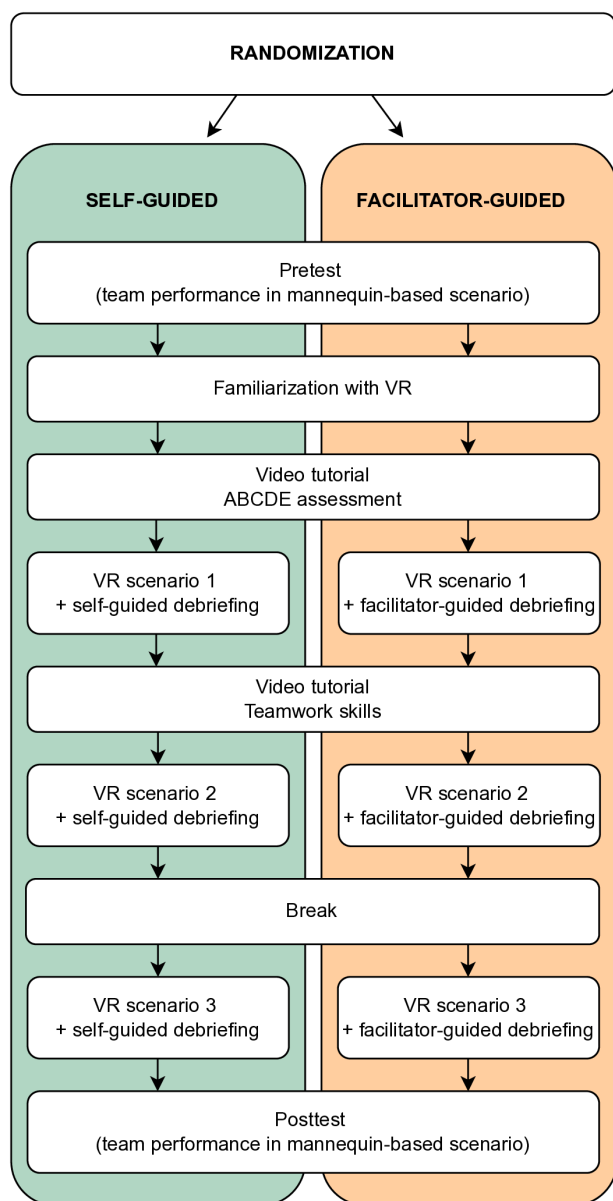
Fig 2. Illustration of pretest, posttest, and intervention phases. The self-guided group will receive a debriefing script based on the Promoting Excellence And Reflective Learning in Simulation (PEARLS) framework [43] adapted for VR and self-guided use [30,44]. No facilitator will be present, but participants may request technical support. A research assistant will introduce the VR simulation, administer questionnaires, and confirm that all scenarios and debriefings were completed.

A.M.S., L.P., J.K., and A.P.; validated by A.P., J.L.S., L.K.G., and S.L.; and implemented in UbiSim's© VR platform by A.M.S. and I.M.H. Scenarios will be delivered using Oculus Quest 2 headsets.

Development of assessment scenarios

The pretest and posttest assessment scenarios depict an 8-year-old with asthma and a 3-month-old with pneumonia. Participants will perform the scenarios in a mannequin-based simulation to ensure that any improvement in team performance is not

due to familiarization with the VR simulation. To reduce potential carryover or learning effects between pre- and post-intervention assessments, distinct but comparable scenarios are used at each time point. Scenarios are matched in complexity and learning objectives but differ in content to minimize practice effects. The two assessment scenarios were tested in our pilot study and found to be similar in difficulty level [30]. A.M.S., J.K. and A.P. created the cases, with revisions from J.L.S., L.K.G., L.P., and S.L.

Outcome measures

Outcome measures and assessment tools are further described in Table 1.

Baseline characteristics will be collected on all participants prior to the intervention, including gender, age, current semester, and prior experience in simulation-based training, computer games, and immersive VR.

RQ1- effect on teamwork skills (primary outcome). Two independent raters blinded to group allocation will rate teamwork skills in the pretest and posttest videos using Clinical Teamwork Scale (CTS) [45]. CTS is designed and validated to rate teamwork skills in simulated pediatric emergencies [45]. CTS has been professionally translated to Danish and demonstrated excellent interrater reliability in our pilot study [30].

RQ2 – effect on overall team performance (secondary outcomes). Two independent raters blinded to group allocation will rate team performance in pretest and posttest videos using the ABCDE checklist, time to critical actions, and CTS.

The ABCDE checklist is modified from the checklist by Hultin et al [46], and measures adherence to the European Pediatric Advanced Life Support guidelines on ABCDE assessment [30,52]. The checklist has demonstrated high interrater reliability in our pilot study [30].

Time to critical actions will be measured as time to each of the following predefined actions: administration of oxygen (pneumonia case) or beta2 agonist (asthma case), administration of fluid bolus 10 mL/kg body weight, and completion of ABCDE assessment.

RQ3 – perceived effectiveness and engagement (secondary outcomes). Participants will complete the Debriefing Assessment for Simulation in Healthcare Student Version (DASH-SV) [47] and Intrinsic Motivation Inventory (IMI) [48].

DASH-SV is a validated questionnaire designed for learners to evaluate the effectiveness of a debriefing. For the self-guided group, 'The instructor' was replaced with 'We' in consultation with the DASH-SV© developers to reflect the self-guided context, e.g., replacing "The instructor provoked in-depth discussions that led me to reflect on my performance" with "We provoked in-depth discussions that led me to reflect on my performance".

IMI is a validated measure of intrinsic motivation related to a specific activity [48]. In this study, we will use the 7-item interest/enjoyment subscale, which is considered the direct measure of intrinsic motivation. IMI and DASH-SV have previously been translated professionally to Danish and tested in a Danish context [30,53].

RQ4 – cognitive load (secondary outcomes). Participants will complete the validated NASA Task Load Index (NASA-TLX) [49], a tool designed to assess perceived workload. A professional Danish translation has previously been completed [30].

RQ5 – cybersickness and usability (secondary outcomes). Participants will report symptoms of cybersickness using the validated Virtual Reality Sickness Questionnaire (VRSQ) [51], and perceived usability will be assessed with the validated System Usability Scale (SUS) [50]. SUS has been professionally translated and validated in a Danish context [48]. At the bottom of the VRSQ, we will add a field asking participants to indicate the duration of their symptoms.

Data collection and management

Participant recruitment will take place from May 6th, 2025, to October 31st, 2025. Data collection began on May 6th, 2025, and is expected to be completed by November 2025. Results are anticipated by June 2026. Schedule of enrolment, interventions, and assessments is presented in Fig 1.

Table 1. Assessment tools, research hypothesis, and outcome measures.

	Assessment tool	Research question	Outcome measure
Team-level outcome (n = 22 teams in each arm)	**Primary outcome**		
	Clinical Teamwork Scale (CTS®) [45] − 15 teamwork skills in five main domains: overall teamwork, communication, situational awareness, decision-making − 11-point Likert scale ranging from 1 = unacceptable to 11 = perfect	RQ1: Is the change from pretest to posttest in the self-guided group non-inferior compared to the facilitator-guided group, with a non-inferiority margin of 0.5? RQ3: Is there a significant increase from pretest to posttest in both groups?	Mean CTS score (mean across 15 items and across the two raters.)
	Secondary outcomes		
	ABCDE checklist [46] − 20 items related to paediatric ABCDE assessment - Checklist ranges from 0 = not initiated to 3 = performed consistently during the whole simulation	RQ2: Is the change from pretest to posttest in the self-guided group comparable to that in the facilitator-guided group? RQ3: Is there a significant increase from pretest to posttest in both groups?	Mean ABCDE checklist score (mean across 20 items and across the two raters.)
	Time to critical actions Time (in seconds) to three predefined actions. If the team does not perform the action, they will be assigned a time of 600 seconds (maximum length of scenario).	RQ2: Is the change from pretest to posttest in the self-guided group comparable to that in the facilitator-guided group? RQ3: Is there a significant increase from pretest to posttest in both groups?	Mean time (mean across the three predefined critical actions and across the two raters.)
Individual level outcomes (n = 44 participants in each arm)	**Debriefing Assessment for Simulation in Healthcare Student Version (DASH-SV)** [47] − 6 items related to perceived effectiveness of the debriefing − 7-point Likert scale ranging from 1 = extremely ineffective to 7 = extremely effective	RQ4: Is the mean in the self-guided group comparable to that in the facilitator-guided group?	Mean DASH-SV score (mean across six items and across the three debriefings)
	Intrinsic Motivation Inventory (IMI) [48] − 7 items from the interest/enjoyment dimension, related to participants motivation. − 7-point Likert scale ranging from 1 = not at all true to 7 = very true	RQ4: Is the mean in the self-guided group comparable to that in the facilitator-guided group?	Mean IMI interest/enjoyment score (mean across seven items)
	NASA Task Load Index (NASA-TLX) [49]. - Six items related to participants perceived workload. - Scale ranging from 0 to 21. Scores are subsequently converted into a single score from 0-100.	RQ5: Is the mean in the self-guided group comparable to that in the facilitator-guided group?	Mean NASA-TLX score (mean across six items)
	System Usability Scale (SUS) [50] − 10 items related to perceived usability. - Likert-scale ranging from 1 = strongly disagree, 5 = strongly agree. The scores are subsequently converted into a single SUS score from 0-100 [50].	Exploratory – descriptive statistics	Mean SUS score (mean across ten items)
	Virtual Reality Sickness Questionnaire (VRSQ) [51] − 9 symptoms of cybersickness - Likert-type scale ranging from 0 = none to 3 = severe.	Exploratory – descriptive statistics	Mean VRSQ score (mean across 9 items)

ABCDE, airways, breathing, circulation, disability, and exposure; NASA, national aeronautics and space administration task load index; RQ, research questions.

Teams will be assessed before the intervention managing a mannequin-based pediatric emergency (pretest) and again after the intervention (posttest). The mannequin-based simulation will be video recorded and two independent, blinded raters will rate team performance based on the videos. Teams will be randomly assigned one case at pretest and complete

the other at posttest. Scenarios will be recorded using a GoPro HERO 4 camera. The raters will be a medical doctor and a medical student, both trained and experienced in assessing team performance using these instruments.

Additionally, after the intervention all participants will complete questionnaires on perceived debriefing quality, motivation, workload, cybersickness, and usability (Table 1 and Fig 1).

Pretest and posttest videos will be stored on a secure hospital drive as logged files, accessible to only to authorized study personnel and are permanently deleted after completion of analyses. All other data will be stored in a secure online database (REDCap) [54] and be pseudonymized before imported into R (version 4.1.2) for statistical analysis [55].

Sample size calculation

The sample size calculation is based on the assumption that self-guided debriefing is non-inferior to facilitator-guided debriefing for the primary outcome (teamwork skills measured by the CTS© scale.) We set a non-inferiority margin of 0.5 points on the 0–10 CTS© scale, meaning that a difference of less than 0.5 points on the CTS© scale is not expected to constitute a meaningful difference in teamwork skills. The non-inferiority margin was informed by expert consensus, our experience rating teamwork skills in the pilot study, and pilot data [30]. Based on a one-sided two-sample t-test with a non-inferiority margin of 0.5, an alpha of 0.05, a power of 80%, and a standard deviation of 0.63 derived from our pilot data [30], the required sample size is 20 teams (40 participants) in each arm. Assuming a 10% dropout rate, we plan to include approximately 44 participants in each arm, and 88 participants in total. The sample size calculation was performed in R version 4.1.2 [55].

Randomization and blinding

Medical students will enroll in the study in teams of two based on schedule availability, without matching for teamwork experience, clinical rotation level, or other characteristics. Each team will be treated as a cluster and randomly allocated to the self-guided or facilitator-guided group at a 1:1 allocation ratio.

The principal investigator will generate the randomization sequence using Sealed Envelope's online randomization service [56]. Upon participant enrolment, the principal investigator will assign each team a team-ID based on the order of enrolment and allocate teams to interventions according to the randomization sequence.

The raters assessing the pretest and posttest videos will be blinded to group allocation. Due to the nature of the interventions, blinding of participants is not feasible. However, participants will not be informed of their group allocation until the intervention day, after completing the pretest, the VR scenarios used for intervention delivery are pre-programmed and identical across the intervention groups to ensure standardization, and the performance evaluation scenarios will follow standardized protocols. The researchers doing the statistical analyses will be blinded to allocation of participants.

Statistical analysis

We will report descriptive statistics on baseline characteristics of each group, reporting continuous variables using median and interquartile range and categorical variables as n/total N (%). Descriptive statistics will be reported for all outcomes at each time point as mean, standard deviations, median, and interquartile range.

We will analyze the team-level data (RQ1 and RQ2) using a constrained linear mixed model with inherent baseline adjustment. Visit (pretest or posttest), assessment scenario (pneumonia or asthma), and the constrained interaction between visit and treatment group (self-guided vs facilitator-guided) will be included as fixed effects. An unstructured covariance pattern will be applied to account for the correlation between repeated measurements and potential changes in variance over time and account for missing data. The mean between the two raters will be calculated prior to statistical analysis.

To analyze individual-level outcomes (RQ3 and RQ4), we will apply a constrained linear mixed model with treatment group as fixed effect and a blocked compound symmetry pattern to account for intra-class correlation between individuals in the same team.

Missing data will be handled within the linear mixed model under the assumption that data are missing at random by including all available data points without requiring imputation. For missing individual-level covariate or questionnaire data, we will first assess the extent and pattern of missingness. If missingness exceeds a minimal threshold of >5%, we will apply multiple imputation using chained equations to account for missing values [57].

For the primary outcome (RQ1), we will report point estimates and two-sided 95% CI. Non-inferiority will be concluded, and the null hypothesis rejected, if the lower bound of the CI exceeds the negative non-inferiority margin [58,59].

For secondary outcomes (RQ2, RQ3, and RQ4), we will report point estimates, two-sided 95% CI, and p-values, and the null hypothesis will be rejected if $p < 0.05$. RQ5 related outcomes will be reported descriptively using median and inter-quartile range.

We will adjust for multiple testing using the Holm method to control the family-wise error rate within each family of research questions, except for the primary outcome [58,60].

Ethics and dissemination

This study is approved by the Danish Data Protection Agency (Privacy, P-2023–167), and the Danish National Committee on Health Research Ethics granted an exemption from requiring ethical approval (F-23005502). The study will adhere to the Declaration of Helsinki.

Participation is voluntary, and the principal investigator will collect written informed consent and permission to record and store video data for the research project from all participants upon enrolment (see supplementary material 1.) Participants may withdraw at any time without consequence, and the intervention will be discontinued if consent is revoked. Data will be stored securely in REDCap [54], with access limited to study investigators. Video recordings are stored on secure, access-restricted institutional servers, accessible only to authorized study personnel. All recordings are used solely for blinded performance assessment and are permanently deleted after completion of the analyses in accordance with the data management plan approved by the Danish Data Protection Agency.

The trial is registered on ClinicalTrials.gov, identifier: NCT06956833, https://clinicaltrials.gov/study/NCT06956833?term=NCT06956833&rank=1 on 04/24/2025. Any protocol changes will be updated on the registry. Results from the study will be submitted to peer-reviewed open-access journals and presented at national and international conferences. Participants who request it will receive a summary of the trial findings.

Discussion

This study is expected to provide valuable insight into the effects of self-guided versus facilitator-guided debriefing in the context of immersive VR. The findings are expected to inform the future development and integration of VR into simulation-based medical education, expanding the understanding of both the potential advantages and limitations of self-guided debriefing.

Findings from our pilot study suggest that both debriefing methods are acceptable, usable, and effective for training interprofessional teams of medical doctors and nurses in a high resource setting [30]. We will add to these findings by investigating the two debriefing approaches in a population of more novice learners. In the absence of a facilitator, learners with less developed self-regulated learning skills may struggle to frame their experiences effectively [27,61], miss critical learning points, and experience a high cognitive load [27,28].

Further, our pilot study was based on a convenience sample. This study will add to our pilot findings by applying a non-inferiority RCT design, thus enabling us to investigate whether self-guided debriefing can be considered non-inferior to facilitator-guided debriefing for improving medical students' teamwork skills.

If self-guided debriefing proves comparably effective to facilitator-guided debriefing, it could enable greater flexibility in the timing, duration, and location of simulation-based medical education, potentially increasing access to simulation-based education in a global perspective. However, several barriers to implementing VR-based simulation remain, such as

logistical, cultural, and technological barriers [11,30,44,62]. Further, this trial involves medical students in a high-income country with access to advanced technology and little is known about the feasibility and effect of VR-based simulation in low resource settings [41,63,64]. Future studies could explore the feasibility and effect of VR in other contexts, particularly those with limited access to advanced technology, technical support, or prior simulation experience, to further inform for whom and under what conditions it is most effective.

Strengths and limitations

Key strengths of this study include its randomized controlled design and the assessment of team performance by blinded raters using video recordings of participants. Further, we will employ a broad range of validated assessment tools, professionally translated to Danish and assessed as feasible for use in a Danish context. The preceding pilot study informed recruitment strategies, intervention delivery, and assessment tool feasibility, strengthening the trial's methodological foundation.

However, several limitations should be noted. First, the participant sample consists of medical students rather than residents or practicing clinicians, which may limit ecological validity and the direct applicability of findings to clinical practice. Second, the trial evaluates outcomes only in the short term, without long-term follow-up to assess retention of knowledge, maintenance of teamwork skills, or transfer of learning to real clinical environments. However, this design will help ensure a robust dataset with minimal missing data. Third, due to the inherent nature of comparing self-guided with facilitator-guided debriefing, participants cannot be blinded to group allocation, which may introduce performance bias. We mitigate this through standardized scenarios for both intervention delivery and performance evaluation, blinding of outcome assessors, use of validated assessment tools, and randomization to ensure balanced groups.

Fourth, randomization was performed at the team (cluster) level without stratification for potential confounding variables (e.g., prior VR or simulation experience), and team composition was not matched for, e.g., prior teamwork experience or clinical rotation level, which may introduce variability in baseline teamwork performance between groups. While baseline characteristics will be compared between groups, chance imbalances cannot be entirely excluded. Fifth, repeated testing introduces the risk of carryover or practice effects. Although scenarios differ in content but are matched for complexity, repeated exposure may lead to learning independent of the intervention. However, this is likely to affect both groups equally, potentially exaggerating overall improvements but not the primary non-inferiority comparison. Finally, although self-guided debriefing will be supported by standardized prompts to ensure consistency, there remains a possibility of variability in the depth and quality of reflection achieved between individuals and teams. These factors should be considered when interpreting the results.

Supporting information

S1 File. SPIRIT checklist 2025.
(PDF)

S2 File. Synopsis of protocol for ethical review inquiry (Danish Original).
(PDF)

S3 File. Synopsis of protocol for ethical review inquiry (English translation).
(PDF)

Acknowledgments

Author contributions

Conceptualization: Amalie Middelboe Sohlin, Anja Poulsen, Jette Led Sørensen, Jesper Kjærgaard.

Funding acquisition: Amalie Middelboe Sohlin, Anja Poulsen, Jette Led Sørensen, Jesper Kjærgaard.

Methodology: Amalie Middelboe Sohlin, Anja Poulsen, Ida Madeline Hoffmann, Line Klingen Gjærde, Stine Lund, Gritt Overbeck, Lone Paulsen, Todd P. Chang, Joy Yeonjoo Lee, Jette Led Sørensen, Jesper Kjærgaard.

Project administration: Amalie Middelboe Sohlin, Anja Poulsen, Jette Led Sørensen, Jesper Kjærgaard.

Supervision: Anja Poulsen, Jette Led Sørensen, Jesper Kjærgaard.

Validation: Anja Poulsen, Line Klingen Gjærde, Stine Lund, Gritt Overbeck, Todd P. Chang, Joy Yeonjoo Lee, Jette Led Sørensen, Jesper Kjærgaard.

Writing – original draft: Amalie Middelboe Sohlin.

Writing – review & editing: Amalie Middelboe Sohlin, Anja Poulsen, Ida Madeline Hoffmann, Line Klingen Gjærde, Stine Lund, Gritt Overbeck, Lone Paulsen, Todd P. Chang, Joy Yeonjoo Lee, Jette Led Sørensen, Jesper Kjærgaard.

References

1. Thim S, Henriksen TB, Laursen H, Schram AL, Paltved C, Lindhard MS. Simulation-Based Emergency Team Training in Pediatrics: A Systematic Review. Pediatrics. 2022;149(4):e2021054305. https://doi.org/10.1542/peds.2021-054305 PMID: 35237809

2. Cheng A, Lang TR, Starr SR, Pusic M, Cook DA. Technology-enhanced simulation and pediatric education: a meta-analysis. Pediatrics. 2014;133(5):e1313-23. https://doi.org/10.1542/peds.2013-2139 PMID: 24733867

3. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. JAMA - J Am Med Assoc. 2011;306(9):978–88.

4. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Acad Med. 2011;86(6):706–11. https://doi.org/10.1097/ACM.0b013e318217e119 PMID: 21512370

5. Cook DA, Brydges R, Hamstra SJ, Zendejas B, Szostek JH, Wang AT, et al. Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis. Simul Healthc. 2012;7(5):308–20. https://doi.org/10.1097/SIH.0b013e3182614f95 PMID: 23032751

6. Zendejas B, Brydges R, Wang AT, Cook DA. Patient outcomes in simulation-based medical education: a systematic review. J Gen Intern Med. 2013;28(8):1078–89.

7. Cook DA, Andersen DK, Combes JR, Feldman DL, Sachdeva AK. The value proposition of simulation-based education. Surgery. 2018;163(4):944–9. https://doi.org/10.1016/j.surg.2017.11.008 PMID: 29452702

8. Foronda CL, Gonzalez L, Meese MM, Slamon N, Baluyot M, Lee J, et al. A Comparison of Virtual Reality to Traditional Simulation in Health Professions Education: A Systematic Review. Simul Healthc. 2024;19(1S):S90–7. https://doi.org/10.1097/SIH.0000000000000745 PMID: 37651101

9. Abbas JR, Chu MMH, Jeyarajah C, Isba R, Payton A, McGrath B, et al. Virtual reality in simulation-based emergency skills training: A systematic review with a narrative synthesis. Resusc Plus. 2023;16:100484. https://doi.org/10.1016/j.resplu.2023.100484 PMID: 37920857

10. Greif R, Bray JE, Djärv T, Drennan IR. International consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Circulation. 2024;145:645–721.

11. Stefanidis D, Cook D, Kalantar-Motamedi S-M, Muret-Wagstaff S, Calhoun AW, Lauridsen KG, et al. Society for Simulation in Healthcare Guidelines for Simulation Training. Simul Healthc. 2024;19(1S):S4–22. https://doi.org/10.1097/SIH.0000000000000776 PMID: 38240614

12. McGrath JL, Taekman JM, Dev P, Danforth DR, Mohan D, Kman N, et al. Using Virtual Reality Simulation Environments to Assess Competence for Emergency Medicine Learners. Acad Emerg Med. 2018;25(2):186–95. https://doi.org/10.1111/acem.13308 PMID: 28888070

13. INACSL Standards Committee. INACSL Standards of Best Practice: SimulationSM Debriefing. Clin Simul Nurs. 2016;12:S21–5. https://doi.org/10.1016/j.ecns.2016.09.008

14. Sawyer T, Eppich W, Brett-Fleegler M, Grant V, Cheng A. More Than One Way to Debrief: A Critical Review of Healthcare Simulation Debriefing Methods. Simul Healthc. 2016;11(3):209–17. https://doi.org/10.1097/SIH.0000000000000148 PMID: 27254527

15. Duff JP, Morse KJ, Seelandt J, Gross IT, Lydston M, Sargeant J, et al. Debriefing Methods for Simulation in Healthcare: A Systematic Review. Simul Healthc. 2024;19(1S):S112–21. https://doi.org/10.1097/SIH.0000000000000765 PMID: 38240623

16. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach. 2005;27(1):10–28. https://doi.org/10.1080/01421590500046924 PMID: 16147767

17. Cheng A, Eppich W, Grant V, Sherbino J, Zendejas B, Cook DA. Debriefing for technology-enhanced simulation: a systematic review and meta-analysis. Med Educ. 2014;48(7):657–66. https://doi.org/10.1111/medu.12432 PMID: 24909527

18. Keiser NL, Arthur W. A meta-analysis of the effectiveness of the after-action review (or debrief) and factors that influence its effectiveness. J Appl Psychol. 2021;106(7):1007–32. https://doi.org/10.1037/apl0000821 PMID: 32852990

19. Tannenbaum SI, Cerasoli CP. Do team and individual debriefs enhance performance? A meta-analysis. Hum Factors. 2013;55(1):231–45. https://doi.org/10.1177/0018720812448394 PMID: 23516804

20. Kolbe M, Schmutz S, Seelandt JC, Eppich WJ, Schmutz JB. Team debriefings in healthcare: aligning intention and impact. BMJ. 2021;374:1–5.

21. Boet S, Bould MD, Sharma B, Revees S, Naik VN, Triby E, et al. Within-team debriefing versus instructor-led debriefing for simulation-based education: a randomized controlled trial. Ann Surg. 2013;258(1):53–8. https://doi.org/10.1097/SLA.0b013e31829659e4 PMID: 23728281

22. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Routledge Handbook of Adapted Physical Education. 2000;55(1):296–312.

23. Eddy ER, Tannenbaum SI, Mathieu JE. Helping teams to help themselves: Comparing two team-led debriefing methods. Pers Psychol. 2013;66(4):975–1008.

24. Verkuyl M, Lapum JL, Hughes M, McCulloch T, Liu L, Mastrilli P, et al. Virtual Gaming Simulation: Exploring Self-Debriefing, Virtual Debriefing, and In-person Debriefing. Clin Simul Nurs. 2018;20:7–14. https://doi.org/10.1016/j.ecns.2018.04.006

25. Strandbygaard J, Bjerrum F, Maagaard M, Winkel P, Larsen CR, Ringsted C, et al. Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial. Ann Surg. 2013;257(5):839–44. https://doi.org/10.1097/SLA.0b013e31827eee6e PMID: 23295321

26. Aagesen AH, Jensen RD, Cheung JJH, Christensen JB, Konge L, Brydges R, et al. The Benefits of Tying Yourself in Knots: Unraveling the Learning Mechanisms of Guided Discovery Learning in an Open Surgical Skills Course. Acad Med. 2020;95(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 59th Annual Research in Medical Education Presentations):S37–43. https://doi.org/10.1097/ACM.0000000000003646 PMID: 32769466

27. Sandars J, Cleary TJ. Self-regulation theory: applications to medical education: AMEE guide no. 58. Med Teach. 2011;33(11):875–86.

28. Seufert T. The interplay between self-regulation in learning and cognitive load. Educ Res Rev. 2018;24:116–29. https://doi.org/10.1016/j.edurev.2018.03.004

29. MacKenna V, Díaz DA, Chase SK, Boden CJ, Loerzel V. Self-debriefing in healthcare simulation: An integrative literature review. Nurse Educ Today. 2021;102:104907. https://doi.org/10.1016/j.nedt.2021.104907 PMID: 33901867

30. Sohlin AM, Hoffman IM, Poulsen A, Gjærde LK, Chang TP, Lee JY, et al. Facilitator-guided vs Self-guided debriefing in Immersive Virtual Reality Paediatric Emergency Team Training: A Randomised Pilot Study on Learning Outcomes and Feasibility. Work in progress; Submitted to European Journal of Pediatrics. 2025;

31. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System. Nursing Critical Care. 2000.

32. WHO. Topic 2: What is human factors and why is it important to patient safety? 2021. Available from: https://www.who.int

33. Abildgren L, Lebahn-Hadidi M, Mogensen CB, Toft P, Nielsen AB, Frandsen TF, et al. The effectiveness of improving healthcare teams' human factor skills using simulation-based training: a systematic review. Adv Simul (Lond). 2022;7(1):12. https://doi.org/10.1186/s41077-022-00207-2 PMID: 35526061

34. Zackoff MW, Real FJ, Sahay RD, Fei L, Guiot A, Lehmann C, et al. Impact of an Immersive Virtual Reality Curriculum on Medical Students' Clinical Assessment of Infants with Respiratory Distress. Pediatr Crit Care Med. 2020;:477–85.

35. Abulfaraj MM, Jeffers JM, Tackett S, Chang T. Virtual reality vs. high-fidelity mannequin-based simulation: a pilot randomized trial evaluating learner performance. Cureus. 2021;13(8).

36. Farra S, Hodgson E, Miller ET, Timm N, Brady W, Gneuhs M, et al. Evacuation of Neonates. Disaster Med Public Health Prep. 2020;13(2):301–8.

37. Katz D, Shah R, Kim E, Park C, Shah A, Levine A, et al. Utilization of a voice-based virtual reality advanced cardiac life support team leader refresher: prospective observational study. J Med Internet Res. 2020;22(3).

38. Issleib M, Kromer A, Pinnschmidt HO, Süss-Havemann C, Kubitz JC. Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students: a randomized controlled trial. Scand J Trauma Resusc Emerg Med. 2021;29(1):1–9.

39. Ferrandini Price M, Escribano Tortosa D, Nieto Fernandez-Pacheco A, Perez Alonso N, Cerón Madrigal JJ, Melendreras-Ruiz R, et al. Comparative study of a simulated incident with multiple victims and immersive virtual reality. Nurse Educ Today. 2018;71:48–53. https://doi.org/10.1016/j.nedt.2018.09.006 PMID: 30241022

40. Berg H, Steinsbekk A. The effect of self-practicing systematic clinical observations in a multiplayer, immersive, interactive virtual reality application versus physical equipment: a randomized controlled trial. Adv Health Sci Educ Theory Pract. 2021;26(2):667–82. https://doi.org/10.1007/s10459-020-10019-6 PMID: 33511505

41. Umoren R, Bucher S, Hippe DS, Ezenwa BN, Fajolu IB, Okwako FM, et al. eHBB: a randomised controlled trial of virtual reality or video for neonatal resuscitation refresher training in healthcare workers in resource-scarce settings. BMJ Open. 2021;11(8):e048506. https://doi.org/10.1136/bmjopen-2020-048506 PMID: 34433598

42. Chan AW, Boutron I, Hopewell S, Moher D, Schulz KF, Collins GS, et al. SPIRIT 2025 Statement: Updated Guideline for Protocols of Randomized Trials. JAMA. 2025.

43. Eppich W, Cheng A. Promoting Excellence and Reflective Learning in Simulation (PEARLS): development and rationale for a blended approach to health care simulation debriefing. Simul Healthc. 2015;10(2):106–15. https://doi.org/10.1097/SIH.0000000000000072 PMID: 25710312

44. Sohlin AM, Kjærgaard J, Hoffman IM, Chang TP, Poulsen A, Lee JY, et al. Immersive virtual reality training: Addressing challenges and unlocking potentials. Med Educ. 2025;:1–13.

45. Guise J-M, Deering SH, Kanki BG, Osterweil P, Li H, Mori M, et al. Validation of a tool to measure and promote clinical teamwork. Simul Healthc. 2008;3(4):217–23. https://doi.org/10.1097/SIH.0b013e31816fdd0a PMID: 19088666

46. Hultin M, Jonsson K, Härgestam M, Lindkvist M, Brulin C. Reliability of instruments that measure situation awareness, team performance and task performance in a simulation setting with medical students. BMJ Open. 2019;9(9):e029412. https://doi.org/10.1136/bmjopen-2019-029412 PMID: 31515425

47. Brett-Fleegler M, Rudolph J, Eppich W, Monuteaux M, Fleegler E, Cheng A, et al. Debriefing assessment for simulation in healthcare: development and psychometric properties. Simul Healthc. 2012;7(5):288–94. https://doi.org/10.1097/SIH.0b013e3182620228 PMID: 22902606

48. Hvidt JCS, Christensen LF, Sibbersen C, Helweg-Jørgensen S, Hansen JP, Lichtenstein MB. Translation and Validation of the System Usability Scale in a Danish Mental Health Setting Using Digital Technologies in Treatment Interventions. Int J Hum–Comput Interact. 2019;36(8):709–16. https://doi.org/10.1080/10447318.2019.1680922

49. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology. 1988;52(C):139–83.

50. Brooke J. SUS: A "quick and dirty" usability scale. Usability evaluation in industry. 2020. p. 207–12.

51. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Appl Ergon. 2018;69:66–73. https://doi.org/10.1016/j.apergo.2017.12.016 PMID: 29477332

52. Van de Voorde P, Turner NM, Djakow J, de Lucas N, Martinez-Mejias A, Biarent D, et al. European Resuscitation Council Guidelines 2021: Paediatric Life Support. Resuscitation. 2021;161:327–87. https://doi.org/10.1016/j.resuscitation.2021.02.015 PMID: 33773830

53. Sørensen JL, van der Vleuten C, Rosthøj S, Østergaard D, LeBlanc V, Johansen M, et al. Simulation-based multiprofessional obstetric anaesthesia training conducted in situ versus off-site leads to similar individual and team outcomes: a randomised educational trial. BMJ Open. 2015;5(10):e008344. https://doi.org/10.1136/bmjopen-2015-008344 PMID: 26443654

54. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377–81.

55. Core Team R. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.r-project.org/

56. https://www.sealedenvelope.com/

57. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1–67.

58. Wiens BL. Multiple comparisons in non-inferiority trials: Reaction to recent regulatory guidance on multiple endpoints in clinical trials. J Biopharm Stat. 2018;28(1):52–62. https://doi.org/10.1080/10543406.2017.1378667 PMID: 29065276

59. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. Trials. 2011;12:106. https://doi.org/10.1186/1745-6215-12-106 PMID: 21539749

60. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. Am J Public Health. 1996;86(5):726–8. https://doi.org/10.2105/ajph.86.5.726 PMID: 8629727

61. Panadero E. A Review of Self-regulated Learning: Six Models and Four Directions for Research. Front Psychol. 2017;8:422. https://doi.org/10.3389/fpsyg.2017.00422 PMID: 28503157

62. Savino S, Mormando G, Saia G, Da Dalt L, Chang TP, Bressan S. SIMPEDVR: using VR in teaching pediatric emergencies to undergraduate students-a pilot study. Eur J Pediatr. 2024;183(1):499–502. https://doi.org/10.1007/s00431-023-05254-z PMID: 37843614

63. Hoffmann IM, Andersen AM, Lund S, Nygaard U, Joshua D, Poulsen A. Smartphone apps hold promise for neonatal emergency care in low-resource settings. Acta Paediatr. 2024;113(12):2526–33. https://doi.org/10.1111/apa.17410 PMID: 39222003

64. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. JMIR Serious Games. 2021;9(3):e29080. https://doi.org/10.2196/29080 PMID: 34255668