PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

Title (Provisional)

Effectiveness of a Virtual Reality-Based Sensory Stimulation Intervention in Preventing Delirium in Intensive Care Units: A Randomized Controlled Trial Protocol

Authors

Liang, Surui; Liu, Yong; Wen, Taoxue; Luo, Dan; he, mingxin; Tian, Jinfei

VERSION 1 - REVIEW

Reviewer	**1**
Name	**von Felten, Stefanie**
Affiliation	**Universitat Zurich, Epidemiology, Biostatistics and Prevention Institute**
Date	**07-Feb-2024**
COI	**I have no competing interests to declare**

Review BMJ Open, Februar 2024

This study protocol describes an RCT on the effectiveness of a virtual reality-based sensory stimulation compared to usual care for the prevention of delirium in intensive care units. The rationale for the study seems to be well argued for, but the study design is not described in full detail. As a consequence, I can not judge at the moment whether it is just the description that is not sufficient or precise enough, or whether the design and methodology suffers from major deficiencies.

Major comments:

- Several primary endpoints without adjusting for multiplicity
- Randomization not clear

My comments are listed in order of relevant SPIRIT items and parts of the manuscript:

Title page

- The protocol version with date and version identifier is missing (SPIRIT item 3)
- Contributors: I cannot connect the abbreviated names with the author names. Please use the Initials to abbreviate authors, for example use SL for Surui Liang or JT for Jinfei Tian.

- The roles and responsibilities of the authors in the trial are not given (SPIRIT item 5a). Who is the principal investigator/co-investigators, who is the trial statistician, who is the sponsor?
- Some of the e-mail addresses of the authors are strange. If available, please provide institutional e-mail addresses (from University or Hospital).

Abstract

- At the beginning, I would expect a sentence or more on the Background/Rationale for the trial.
- BMJ open requires that dates of the study are included in the manuscript. This information (or part of it) may be included in the Abstract (e.g. when the study started and when it is expected to end).
- When reading the Abstract, I noticed that there are several primary endpoints, which made me wondering whether a corresponding adjustment for multiplicity was planned (but see below).
- I was also wondering what a "paired randomization method" was, and how it would practically work (but see below).

Introduction

- Explanation for choice of comparator (SPIRIT item 6b) is missing. If I understand correctly, the comparator is "usual care", but this is not described early in the manuscript.
- P4, lines 8-13: I would expect some references for this "international research" here.
- P5, lines 5-13: You write that VR "can have a protective effect", but give no reference. I therefore assume this should read "may have …" and corresponds to your expectation. Please clarify the sentence.
- P5, lines 38-43: there is something wrong with this sentence
- Aims and hypotheses: a) "A greater" reduction would require information about the comparison. I assume you mean a reduction in incidence compared to the control arm ("greater" is not needed).
  b), c), and e) and p6, line 10-11: I would avoid the word "significant" here, because it is kind of reserved for statistical results that you have in the end.

Methods and Analysis

- Design: I would add "parallel group" after "two-arm". As far as I understand, the trial is done in two different hospitals, so you could add multi-center or bi-center.
- Setting: I would name the two hospitals.
- Sample size determination: the calculation was performed for delirium incidence (binary endpoint), comparing two proportions. Expected proportions are 13% (VR arm) vs. 25% (control arm), significance level 5% and power 80 %. A drop-out rate of 20% was accounted for. Double checking the routine for a McNemar test in G*Power, I saw that an odds ratio is required, which the authors report as 0.57, and a proportion of "discordant pairs" which is missing from the description. Please report the proportion of "discordant pairs" in order to allow the reader to reproduce your calculation. Then, please report the sample size calculated in G*Power (without drop-out) and then the sample size you get if accounting for 20% dropout.
  Moreover, here and for the analyses, I am not totally convinced that it makes sense to use paired tests, because although the randomisation is performed in a paired manner to ensure some balance, the patients that are paired are otherwise not closer to each

other than to the rest of the patients (no relatives for example). I think it would also be ok to use standard tests, here a standard test for comparing two proportions (e.g., a z-test).

- Randomisation: a paired randomisation method is mentioned, matching participants based on a delirium risk scoring model for ICU patients. This method requires a more precise description, especially with regard to the practical implementation. How did you randomize when you include one patient? I assume it was not possible to include patients pairwise, and waiting for a match is also not possible, since the intervention starts upon admission. And how would this work together with consecutively numbered, sealed and opaque envelopes, which would mean that the randomization list is prepared in advance?
  It is usual to use center as a stratification factor or as a variable in a minimization procedure. How do you deal with the two centers in the randomization?
- Intervention: what exactly is a "sample reorientation message"? Is this an information in written form for the patients family or caregivers?
  If family members of patient in the intervention group are informed about delirium and need to provide material for preparing the VR, I would expect that the fact that they know about delirium could change their behaviour in a way that also has a preventive effect on delirium. So in the end, it is not clear whether a possible effect of the intervention is due to VR, the families behaviour or both. If correct, this may be added as a limitation of the study in the Discussion section
  I was further wondering if in practice the families provide the material in time to start the VR intervention on the patient as early as possible. Is there not usually a delay? And if yes, how long?
  How do you intend to deal with medication that may prevent delirium? Is such medication possible? If yes, do you record whether patients receive it and what dose they receive?
- Primary outcomes: It is possible to design a study with more than one primary outcome, but this is more complex than with one primary outcome. For example, some adjustment for multiplicity has to be made, and there are different ways to do that. Since no such thing seems to be planned, I suggest to define just ONE primary outcome. Due to the description of the sample size calculation, the logical choice would be delirium incidence as a binary outcome. More precisely, it would probably be delirium incidence up to discharge or day 14 (if still in hospital then). It is very important to precisely define the primary endpoint with the relevant time period over which it is assessed and how you are using the CAM-ICU score to define delirium incidence.
  Delirium duration and delirium severity (which are only measured if someone has delirium) should be treated as secondary endpoints. They should also be described in a more precise way. For example, is delirium duration the number of shifts the patient is in delirium? And how do you summarize (if you summarize at all) delirium severity over 14 days? Do you use the highest score over the 14 day period?
- Data collection: The principal investigator is usually one person. Can he/she really inform all patients from two hospitals? Or are several persons doing this? And do you have one or several reasearch assistants? These roles should be clarified under item 5a, see above.
- Data management: Excel is mentioned, but Excel is not really GCP conform. How do you manage access to the data and integrity and track record?

- Data analysis, software: there is no R 23.0. Maybe you meant the year of the Version. I am currently using R version 4.3.2 (2023-10-31). You could write R (current version at the time of analysis). Why would you want to use SPSS if you can use R?
- Data analysis, statistical methods: Here, I have the same comment about paired tests as above under sample size calculation. I am not sure they are needed, since the patients are not naturally paired. Further, I assume that mixed-effects models are used for outcomes with repeated measurements. Is this correct? Please specify the models in more detail. Which random terms are used? Which fixed explanatory variables will you use? If you use the same logic as for paired tests, you would need a random intercept for the pair and one for the patient, to account for pairing and repeated measurements from one patients
- Data analysis, missing aspects: are no subgroup and adjusted analyses planned? To cover item 20b, I would recommend to state that, or if actually planned, to add the planned analyses. For example subgroup analyses should be defined a priori in the study protocol. Also, the analysis population should be defined in more detail. Mentioning ITT is not sufficient. And do you plan to analyse imputed data only, or complete cases in addition?
- CONSORT flow diagram: it is great that you provide the flow diagram you intend to present already. Since you assess some things up to day 14 (or discharge) and other things up to 6 months, I would provide information on follow-up for these two follow-up times.

Discussion

- This section is not really a discussion yet. I suggest to place some of its content elsewhere.
- Instead, you could discuss strengths and limitations of your trial

---

Reviewer	**2**
Name	**Kasapoğlu, Elçin**
Affiliation	**Bartin Universitesi**
Date	**20-Feb-2024**
COI	**No**

I am looking forward to the results of your work.

VERSION 1 - AUTHOR RESPONSE

Review BMJ Open, Februar 2024
This study protocol describes an RCT on the effectiveness of a virtual reality-based sensory stimulation compared to usual care for prevention of delirium in intensive care units. The rationale for the study seems to be well argued for, but the study design is not described in full detail. As a consequence, I can not judge at the moment whether it is just the description that is not sufficient or precise enough, or whether the design and methodology suffers from major deficiencies.
Major comments:
- Several primary endpoints without adjusting for multiplicity
- Randomization not clear

Reply:
Primary Endpoints: In response to your comment about adjusting for multiplicity, we have made non-delirium days the main primary outcome and categorized other outcomes as secondary. This adjustment helps to streamline the focus of the study and ensures clarity in our analysis plan.
Randomization Process: We appreciate your observation regarding the clarity of the randomization process. To address this, we have revised the randomization method to ensure transparency and robustness. Specifically, we have implemented separate randomization at each center, removing the paired allocation. However, during the analysis phase, we will consider employing propensity score matching and exploring potential patient factors that may influence the intervention's effectiveness.

My comments are listed in order of relevant SPIRIT items and parts of the manuscript:
Title page
 The protocol version with date and version identifier is missing (SPIRIT item 3)
Contributors: I cannot connect the abbreviated names with the author names. Please use the Initials to abbreviate authors, for example use SL for Surui Liang or JT for Jinfei Tian.
 The roles and responsibilities of the authors in the trial are not given (SPIRIT item 5a). Who is the principal investigator/co-investigators, who is the trial statistician, who is the sponsor?
 Some of the e-mail addresses of the authors are strange. If available, please provide institutional e-mail addresses (from University or Hospital).

Reply:
Protocol Version: We have now included the protocol version with the date and version identifier on the title page, as per SPIRIT item 3.
Contributors: Initials have been used to abbreviate authors, as suggested. Additionally, we have provided the roles and responsibilities of the authors involved in the trial, including the principal investigator, co-investigators, trial statistician, and sponsor, as per SPIRIT item 5a.
Email Addresses: We have replaced the non-standard email addresses with institutional email addresses from our respective universities or institutions.

Abstract
 At the beginning, I would expect a sentence or more on the Background/Rationale for the trial.
 BMJ open requires that dates of the study are included in the manuscript. This information (or part of it) may be included in the Abstract (e.g. when the study started and when it is expected to end).
 When reading the Abstract, I noticed that there are several primary endpoints, which made me wondering whether a corresponding adjustment for multiplicity was planned (but see below).
 I was also wondering what a “paired randomization method” was, and how it would practically work (but see below).

Reply:Background/Rationale: We have updated the background section of the abstract.
Revised Background: "Delirium is a common acute cognitive impairment characterized by confusion, disorientation, and attention deficits, particularly prevalent in ICU settings. Given its significant impact on patients, caregivers, and healthcare resources, preventing delirium in ICU patients is of paramount importance."
Inclusion of Study Dates: We have included information about the study duration in the abstract, specifying that the study will commence in September 2024 and conclude in November 2026, as per BMJ Open requirements.
Multiplicity Adjustment and Paired Randomization: We have addressed the concerns regarding multiplicity adjustment and the paired randomization method in our previous response, as requested.

Introduction
 Explanation for choice of comparator (SPIRIT item 6b) is missing. If I understand correctly, the comparator is “usual care”, but this is not described early in the manuscript.
 P4, lines 8-13: I would expect some references for this “international research” here.
 P5, lines 5-13: You write that VR “can have a protective effect”, but give no reference. I therefore assume this should read “may have …” and corresponds to your expectation.
Please clarify the sentence.

 P5, lines 38-43: there is something wrong with this sentence
 Aims and hypotheses: a) “A greater” reduction would require information about the comparison. I assume you mean a reduction in incidence compared to the control arm (“greater” is not needed). b), c), and e) and p6, line 10-11: I would avoid the word “significant” here, because it is kind of reserved for statistical results that you have in the end.

Reply:
Explanation for choice of comparator: We have clarified early in the manuscript that the comparator is "usual care" and provided an explanation for this choice, as per SPIRIT item 6b.
References for international research: We have reviewed references to support the statement about international research on delirium prevention.
Clarification of VR's protective effect: The sentence has been revised to clarify that VR "may have" a protective effect.
Aims and hypotheses: We have removed the terms "significant" and "greater" to avoid implying statistical significance before the results are obtained.

Design: I would add “parallel group” after “two-arm”. As far as I understand, the trial is done in two different hospitals, so you could add multi-center or bi-center.
Setting: I would name the two hospitals.
Reply:
Design: We have revised it to "A multi-center, assessor-blinded, two-arm parallel group RCT will be conducted."
Setting: We have complemented the two hospitals' names in the setting description.

Sample size determination: the calculation was performed for delirium incidence (binary endpoint), comparing two proportions. Expected proportions are 13% (VR arm) vs. 25% (control arm), significance level 5% and power 80 %. A drop-out rate of 20% was accounted for. Double checking the routine for a McNemar test in G*Power, I saw that an odds ratio is required, which the authors report as 0.57, and a proportion of “discordant pairs” which is missing from the description. Please report the proportion of “discordant pairs” in order to allow the reader to reproduce your calculation. Then, please report the sample size calculated in G*Power (without drop-out) and then the sample size you get if accounting for 20% dropout.
Moreover, here and for the analyses, I am not totally convinced that it makes sense to use paired tests, because although the randomisation is performed in a paired manner to ensure some balance, the patients that are paired are otherwise not closer to each other than to the rest of the patients (no relatives for example). I think it would also be ok to use standard tests, here a standard test for comparing two proportions (e.g., a z-test).

Reply:
We have updated our methodology to employ a standard test for comparing two proportions rather than a paired test. Subsequently, we recalculated the sample size using this standard test in G*Power, and have included the resulting sample size without accounting for dropout.

Randomisation: a paired randomisation method is mentioned, matching participants based on a delirium risk scoring model for ICU patients. This method requires a more precise description, especially with regard to the practical implementation. How did you randomize when you include one patient? I assume it was not possible to include patients pairwise, and waiting for a match is also not possible, since the intervention starts upon admission. And how would this work together with consecutively numbered, sealed and opaque envelopes, which would mean that the randomization list is prepared in advance?
It is usual to use center as a stratification factor or as a variable in a minimization procedure.
How do you deal with the two centers in the randomization?

Reply: We appreciate your observation regarding the clarity of the randomization process. To address this, we have revised the randomization methods. Specifically, we have implemented separate randomization at each center, removing the paired allocation. However, during the analysis phase, we will consider employing propensity score matching and exploring potential patient factors that may influence the intervention's effectiveness.

Intervention: what exactly is a “sample reorientation message”? Is this an information in written form for the patients family or caregivers? If family members of patient in the intervention group are informed about delirium and need to provide material for preparing the VR, I would expect that the fact that they know about delirium could change their behaviour in a way that also has a preventive effect on delirium. So in the end, it is not clear whether a possible effect of the intervention is due to VR, the families behaviour or both. If correct, this may be added as a limitation of the study in the Discussion section

Reply:
A "sample reorientation message" refers to a written communication sample provided to their families or caregivers to help them understand and navigate changes or transitions in care, treatment, or environment.We agree that family involvement may influence patient outcomes and could be considered as a confounding factor in our study. We will carefully consider adding this as a limitation in the discussion section of the manuscript, acknowledging the possibility that observed effects may be due to a combination of the VR intervention and changes in family behavior.

I was further wondering if in practice the families provide the material in time to start the VR intervention on the patient as early as possible. Is there not usually a delay? And if yes, how long?
How do you intend to deal with medication that may prevent delirium? Is such medication possible? If yes, do you record whether patients receive it and what dose they receive?

Reply:We anticipate obtaining materials from family members, such as videos, audio recordings, or photos, at the time of recruitment, which we expect to be promptly available. If there are any delays, they are typically resolved within the following day. We provide contact information and reminders to ensure timely submission of materials. Additionally, we have contingency plans in place to offer standard VR content, such as nature scenes or seascapes, if needed, to ensure timely initiation of the VR intervention.
In our study, our primary focus is on preventing delirium. However, if delirium does occur, physicians may prescribe antipsychotic medications. We will accurately document whether patients receive these medications and the dosage administered. Furthermore, we will maintain detailed records of the dosage of sedatives and analgesics administered to patients throughout the study.

Primary outcomes: It is possible to design a study with more than one primary outcome, but this is more complex than with one primary outcome. For example, some adjustment for multiplicity has to be made, and there are different ways to do that. Since no such thing seems to be planned, I suggest to define just ONE primary outcome. Due to the description of the sample size calculation, the logical choice would be delirium incidence as a binary outcome. More precisely, it would probably be delirium incidence up to discharge or day 14 (if still in hospital then). It is very important to precisely define the primary endpoint with the relevant time period over which it is assessed and how you are using the CAM-ICU score to define delirium incidence.
Delirium duration and delirium severity (which are only measured if someone has delirium) should be treated as secondary endpoints. They should also be described in a more precise way. For example, is delirium duration the number of shifts the patient is in delirium? And how do you summarize (if you summarize at all) delirium severity over 14 days? Do you use the highest score over the 14 day period?

Reply:
Primary Outcome: We defined delirium-free days over a 14-day period as the primary outcome.
Secondary Outcomes:
Delirium Duration: We will measure delirium duration for each shift, and summarize it in days. Delirium Severity: We will use the highest score of the Confusion Assessment Method for the ICU (CAM-ICU) over the 14-day period to represent delirium severity.

Data collection: The principal investigator is usually one person. Can he/she really inform all patients from two hospitals? Or are several persons doing this? And do you have one or several reasearch assistants? These roles should be clarified under item 5a, see above.

Reply: Yes, we have four research assistants who have received consistent training in obtaining written informed consent from participants and conducting data collection procedures. These assistants, under the

supervision of the principal investigator, are tasked with informing patients and collecting data at two study hospitals.

Data management: Excel is mentioned, but Excel is not really GCP conform. How do you manage access to the data and integrity and track record?

Reply: Each patient has a case report form (CRF) for data collection, initially documented on paper. Following this, the data from the CRFs are transferred to Excel spreadsheets on computers. Access to the Excel files will be restricted through password protection. Only authorized personnel participating in the study will have access to these files. All data entry and manipulation activities will be logged and tracked to maintain transparency and accountability. Furthermore, regular audits and quality checks will be conducted to validate the accuracy and consistency of the data.

Data analysis, software: there is no R 23.0. Maybe you meant the year of the Version. I am currently using R version 4.3.2 (2023-10-31). You could write R (current version at the time of analysis). Why would you want to use SPSS if you can use R?
Reply: Yes, we used R version 4.3.2 for data analysis, which was the current version at the time of our analysis. We chose R for its versatility and extensive statistical capabilities. However, since some analysts on our team are more familiar with SPSS, we initially considered using both R and SPSS to leverage their expertise and ensure comprehensive analysis of the data. However, upon further consideration, we have decided to proceed with R exclusively to maintain consistency in our analysis approach.

Data analysis, statistical methods: Here, I have the same comment about paired tests as above under sample size calculation. I am not sure they are needed, since the patients are not naturally paired. Further, I assume that mixed-effects models are used for outcomes with repeated measurements. Is this correct? Please specify the models in more detail. Which random terms are used? Which fixed explanatory variables will you use? If you use the same logic as for paired tests, you would need a random intercept for the pair and one for the patient, to account for pairing and repeated measurements from one patients
Data analysis, missing aspects: are no subgroup and adjusted analyses planned? To cover item 20b, I would recommend to state that, or if actually planned, to add the planned analyses. For example subgroup analyses should be defined a priori in the study protocol. Also, the analysis population should be defined in more detail. Mentioning ITT is not sufficient. And do you plan to analyse imputed data only, or complete cases in addition?

Reply:
Paired tests: Following your recommendation, we have decided to cancel the paired tests. Given that the patients are not naturally paired, we agree that paired tests may not be appropriate in this context.
Statistical models for outcomes with repeated measurements: We will utilize mixed-effects models for outcomes with repeated measurements. Mixed-effects models, also known as multilevel or hierarchical models, are statistical models used for analyzing data that have a hierarchical or nested structure. They are particularly useful when dealing with repeated measurements. Fixed effects are used to estimate the overall population-level relationships between predictors and the outcome. For our study, we plan to include predictors such as APACHE-II score, age, admission category, and use of sedatives and analgesics as fixed explanatory variables.
For statistical analysis, we designate two research personnel with master's degrees. Subgroup analyses will be conducted based on factors such as age, gender, and length of ICU stay.
In terms of data analysis approach, we plan to analyze both imputed data and complete cases. Specifically, we will conduct intention-to-treat (ITT) and per-protocol (PP) analyses to ensure robustness in our findings.

CONSORT flow diagram: it is great that you provide the flow diagram you intend to present already. Since you assess some things up to day 14 (or discharge) and other things up to 6 months, I would provide information on follow-up for these two follow-up times.
Reply:
When disseminating the research results, we will include information on the follow-up durations for both short-term (up to day 14 or discharge) and long-term (up to 6 months) assessments in our CONSORT flow diagram. However, at the current status, we have used the standard CONSORT flow diagram.

Discussion
- This section is not really a discussion yet. I suggest to place some of its content elsewhere.
- Instead, you could discuss strengths and limitations of your trial

Reply:
Yes, we have removed the content that was not appropriate for the discussion section and have focused on discussing the strengths and limitations of our trial. Thank you for your suggestion.

VERSION 2 - AUTHOR RESPONSE

4. Reviewer 1: P5, lines 38-43: there is something wrong with this sentence
Editor: No response was provided to this comment.

New Reply: I have revised the sentence. The updated sentence is as follows:

"Therefore, VR-based visual and auditory stimuli, by immersing patients in relaxing VR environments, may provide a protective effect against environmental stress and support the recovery of physiological, emotional, and attentional functions, thereby helping to prevent delirium."

5.- Please clarify where the following reviewer comments have been addressed in your revised manuscript:

Reviewer 1: P4, lines 8-13: I would expect some references for this “international research” here
Author: References for international research: We have reviewed references to support the statement about international research on delirium prevention
Editor: Please clarify what references were added to support this statement, there does not appear to be any references added to your revised text.

New Reply: I apologize for any confusion in the previous statement, and I have revised the entire sentence for clarity.

“International research (15–17) suggests that VR offers significant advantages in sensory stimulation and could be a promising approach for preventing ICU delirium. However, it is important to note that studies on the use of VR in this context are still in the feasibility testing stage.”

6. Reviewer: Intervention: what exactly is a “sample reorientation message”? Is this an

information in written form for the patients family or caregivers? If family members of patient in the intervention group are informed about delirium and need to provide material for preparing the VR, I would expect that the fact that they know about delirium could change their behaviour in a way that also has a preventive effect on delirium.
So in the end, it is not clear whether a possible effect of the intervention is due to VR, the families behaviour or both. If correct, this may be added as a limitation of the study in the Discussion section
Author Reply: A "sample reorientation message" refers to a written communication sample provided to their families or caregivers to help them understand and navigate changes or transitions in care, treatment, or environment. We agree that family involvement may influence patient outcomes and could be considered as a confounding factor in our study. We will carefully consider adding this
as a limitation in the discussion section of the manuscript, acknowledging the possibility that observed effects may be due to a combination of the VR intervention and changes in family behavior.
Editor: Please include your response to the reviewer in the text of your revised manuscript.

New Reply: Yes, we have added this information to the discussion.

“In the intervention, family caregivers are involved through a "sample reorientation message," which refers to a written communication provided to help them understand and navigate changes in care, treatment, or environment. We acknowledge that family involvement may influence patient outcomes and could act as a confounding factor in our study. The observed effects may result from a combination of the VR intervention and changes in family behavior.”