

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)

Effects of Philips Visual Patient Avatar on vital sign deviations and audible alarm burden in perioperative care: a dual-centre, quasi-experimental pre-post big-data study protocol

Authors

Jiang, Silis Y; Roche, Tadzio R; Cybulski, Krzysztof; Dugac, Gaspar; Meier, Lukas; Tangel, Virginia E; Ebensperger, Max; Maskos, Andreas; Tucci, Michael; Nöthiger, Christoph B.; Kalisch, Markus; Turnbull, Z A; Tscholl, David W.

VERSION 1 - REVIEW

Reviewer	1
Name	Shahidi Delshad, Elham
Affiliation	Tehran University of Medical Sciences, Anesthesia
Date	31-May-2026
COI	None

This protocol addresses an important and timely topic in perioperative monitoring and human factors research. However, several methodological issues require clarification before the protocol can be considered sufficiently robust.

1-The introduction provides a comprehensive overview of avatar-based monitoring technologies; however, much of the supporting evidence originates from studies conducted by the developers of the technology. The authors should incorporate additional independent evidence, if available, to reduce the risk of citation bias and strengthen the rationale for the study.

2-The rationale for selecting vital sign deviation time as the primary outcome requires further justification. The manuscript should explain more explicitly why this process measure is expected to reflect improved situation awareness and clinical performance in real-world settings.

3-The major limitation of this protocol is the uncontrolled pre-post design. Temporal changes unrelated to the intervention (e.g., staffing changes, workflow modifications, institutional

quality-improvement initiatives, changes in case mix, or secular trends) may influence outcomes. The authors should discuss these potential sources of temporal confounding in greater detail and explain whether alternative approaches such as interrupted time-series analysis were considered.

4-A wash-in ("sink-in") phase has been incorporated, which is appropriate. However, additional justification should be provided regarding the selected duration and whether learning effects may continue beyond the predefined adaptation period.

5-The protocol assumes that all post-implementation cases were exposed to the Visual Patient Avatar. However, the supplementary material indicates that clinicians could select different monitor layouts and customize display configurations. Please clarify whether actual usage data, display logs, or layout selection records are available. Without such information, exposure misclassification may occur.

6-The distinction between implementation of the technology and actual utilization of the technology should be discussed more explicitly.

7-Alarm-related outcomes may be influenced by modifications in alarm settings, thresholds, or user preferences over time. The manuscript should clarify whether alarm configurations remained stable throughout the study period and whether any changes in threshold settings will be monitored or adjusted for in the analyses.

8-Because alarm thresholds differ substantially between participating centers, additional discussion is needed regarding the comparability and interpretation of alarm-related outcomes across sites.

9-The statistical analysis plan assumes independence of observations. However, anaesthesia cases are likely clustered within providers, operating rooms, units, and time periods. Ignoring such clustering may lead to underestimated standard errors and inflated statistical significance. The authors should clarify whether provider-level or location-level identifiers are available and whether mixed-effects models or generalized estimating equations were considered.

10-The protocol would benefit from a predefined sensitivity analysis plan. Potential sensitivity analyses could include stratification by surgical urgency, ASA status, procedure complexity, care location, and early versus late post-implementation periods.

11- The manuscript states that Bonferroni correction will be applied within each outcome family. Please provide additional details regarding how hypothesis families will be defined and how many comparisons are anticipated for each outcome domain.

12- The rationale for selecting a 1% missing-data threshold before applying multiple imputation is unclear. Please justify this decision and describe how the missing-data mechanism (MCAR, MAR, or MNAR) will be assessed.

13-The protocol would benefit from a clearer explanation of how adjustment variables were selected. A directed acyclic graph (DAG) or conceptual causal framework could strengthen the justification for the chosen covariates and improve transparency.

14- The discussion appropriately acknowledges several limitations. However, the implications of residual confounding and the inability to establish causal relationships should be emphasized more strongly.

15- Given the substantial financial and intellectual involvement of several authors in the development of the Visual Patient Avatar technology, consideration should be given to describing any measures implemented to ensure analytical independence and minimize potential investigator bias.

16-The term “big-data study” in the title may be unnecessarily promotional. A more neutral term such as “large-scale observational study” may be preferable.

17-The term “sink-in phase” is uncommon in epidemiological and implementation research literature. Consider replacing it with a more widely recognized term such as “adaptation phase” or “wash-in period.”

18- Please indicate whether the protocol follows relevant reporting guidelines for routinely collected health data studies (e.g., RECORD extension of STROBE) and provide the corresponding checklist, if available.

VERSION 1 - AUTHOR RESPONSE

Reviewer 1 comments

We thank Dr Elham Shahidi Delshad for the thoughtful and constructive review. The comments helped us strengthen the protocol substantially.

1. The introduction provides a comprehensive overview of avatar-based monitoring technologies; however, much of the supporting evidence originates from studies conducted by the developers of the technology. The authors should incorporate additional independent evidence, if available, to reduce the risk of citation bias and strengthen the rationale for the study.

Response: We agree and have made the context more explicit. VPA was developed by members of the author group and has only recently entered broader clinical use, which helps explain why independent VPA-specific evidence remains limited. We added independent evidence from the broader human-factors literature on graphical and configural patient-monitoring displays and an independent systematic review of visual patient avatar research. We also state that broader adoption should enable additional independent evaluations and emphasise the present need for transparent prespecification, explicit conflict-of-interest disclosure and analytical safeguards. Location: Introduction p.4, lines 23-37; new references 23-27.

2. The rationale for selecting vital sign deviation time as the primary outcome requires further justification.

Response: We have expanded the 'Rationale for the chosen outcome measures' subsection. The revised text explains that VPA is designed to improve perception and comprehension of vital-sign status; if this translates into earlier corrective action, the most direct measurable signal should be reduced time outside predefined vital-sign ranges. We also clarify that this is an objective, routinely available, case-level process outcome. Location: Rationale for outcomes p.8, lines 20-29.

3. The major limitation of this protocol is the uncontrolled pre-post design. Temporal changes unrelated to the intervention may influence outcomes. The authors should discuss these sources of temporal confounding and explain whether interrupted time-series analysis was considered.

Response: We have substantially revised the rationale, statistical analysis plan and Discussion. The manuscript now explicitly discusses staffing changes, trainee rotations, workflow modifications, quality-improvement activities, seasonal case-mix changes and secular trends. A conventional single-interruption time-series analysis was considered but not selected as the primary design because deployment in large academic hospitals required sequential monitor software updates over several days, followed by training and workflow adaptation. Simultaneous activation across all devices and experimental on-off switching were not feasible in routine care, and the exposure state could not be standardised across every monitor, room and time point. We will describe case mix by phase, adjust for prespecified covariates and conduct sensitivity analyses incorporating calendar time where feasible. We have also strengthened the wording that the study will report associations, not causal effects. Location: Rationale p.5, lines 18-31; Statistical analysis p.10, lines 11-16; Discussion p.11, lines 22-26.

4. A wash-in ('sink-in') phase has been incorporated. Additional justification should be provided regarding the selected duration and whether learning effects may continue beyond the predefined adaptation period.

Response: We have replaced 'sink-in' with 'adaptation phase' throughout the manuscript and added an explicit justification. The approximately six-month adaptation period at each centre was prespecified from the local rollout and training schedules to allow sequential installation, staff exposure and workflow uptake. We do not assume that it eliminates all learning effects; the manuscript now states that learning may continue beyond the formal adaptation phase and prespecifies early-versus-late post-implementation sensitivity analyses. Location: Study design p.5, lines 3-12; timeline and adaptation p.6, lines 1-9; sensitivity analyses p.10, lines 11-16.

5. The protocol assumes that all post-implementation cases were exposed to VPA; however, clinicians could select different monitor layouts and customize display configurations. Please clarify whether actual usage data, display logs or layout selection records are available.

Response: We have added a dedicated 'Implementation versus utilisation' subsection. It clarifies that the exposure is implementation phase rather than individually measured utilisation. VPA is enabled by default and no routine non-VPA monitor layout is available in participating areas. However, the monitors do not routinely retain research-accessible, time-stamped logs of selected layouts or display-state changes. Consequently, we cannot determine which VPA-containing layout was active on a specific monitor at a given time, its relative prominence or clinicians' actual attention. We have added the resulting exposure misclassification as a limitation. Location: Intervention and implementation versus utilisation p.7, lines 1-22; Discussion p.11, lines 27-33.

6. The distinction between implementation of the technology and actual utilization of the technology should be discussed more explicitly.

Response: We have made this distinction explicit in Methods and Discussion. The revised manuscript states that the primary analysis estimates an implementation-level association analogous to an intention-to-treat implementation effect, and that limited clinician attention to the avatar despite availability could dilute observed associations. Location: Implementation versus utilisation p.7, lines 14-22; Discussion p.11, lines 27-33.

7. Alarm-related outcomes may be influenced by modifications in alarm settings, thresholds or user preferences over time. Please clarify whether alarm configurations remained stable and whether changes will be monitored or adjusted for.

Response: We have added an 'Alarm and threshold stability' subsection. It states that centre-specific default alarm settings and VPA thresholds were configured before implementation and kept stable during the prespecified study windows. User-level alarm-limit changes during individual cases are part of routine care and reflected in alarm logs, but they are not uniformly available as analysable covariates across centres. Vital-sign deviation outcomes will use fixed prespecified thresholds, while audible alarm outcomes will use recorded alarm events. Location: Alarm and threshold stability p.7, lines 23-29; Table 2 p.17.

8. Because alarm thresholds differ substantially between participating centres, additional discussion is needed regarding comparability and interpretation of alarm-related outcomes across sites.

Response: We have clarified that analyses will be performed separately by centre and that no pooled cross-centre estimate of absolute alarm burden is planned. The revised Discussion explains that centre-specific alarm configurations limit direct comparison of absolute alarm rates while allowing evaluation of implementation under heterogeneous real-world conditions. Location: Strengths box p.3, lines 35-36; Statistical analysis p.9, lines 12-15; Discussion p.11, lines 17-21.

9. The statistical analysis plan assumes independence of observations. Anaesthesia cases are likely clustered within providers, operating rooms, units and time periods. Please clarify whether identifiers are available and whether mixed-effects models or GEEs were considered.

Response: We have added a paragraph to Statistical analysis. Provider-level identifiers are not available in the anonymised analysis datasets and would not fully represent team-based anaesthesia care; therefore provider-level mixed-effects models or provider-level GEEs are not planned. Care-area identifiers will be used for adjustment or stratification, and location-level fixed effects will be explored where sufficiently complete identifiers are available. We explicitly list residual clustering as a limitation. Location: Statistical analysis p.9, lines 33-38; Discussion p.11, lines 34-37.

10. The protocol would benefit from a predefined sensitivity analysis plan.

Response: We have added prespecified sensitivity analyses, including stratification or interaction analyses by care area, surgical urgency, ASA physical status, surgical risk classification, type of anaesthesia and early-versus-late post-implementation periods. We also added potential sensitivity analyses for alternative outcome distributions, missingness, calendar month or season and location-level fixed effects where data permit. Location: Statistical analysis p.10, lines 11-16.

11. The manuscript states that Bonferroni correction will be applied within each outcome family. Please provide details regarding hypothesis families and anticipated comparisons.

Response: We have clarified that the five outcome families are vital-sign deviation time, VitalSign AUC, alarm count, alarm time and alarm reaction time. Within each family and centre, each analysable vital sign contributes one primary phase-effect p value. The manuscript now states that the

maximum number of comparisons is 15 vital signs at NYP/WCM and 13 at USZ for vital-sign deviation and AUC outcomes, with fewer comparisons for alarm-related outcomes. Location: Statistical analysis p.9, lines 39-41 and p.10, lines 1-3.

12. The rationale for selecting a 1% missing-data threshold before multiple imputation is unclear. Please justify this decision and describe how MCAR, MAR or MNAR will be assessed.

Response: We have revised the missing-data plan. The threshold was changed from 1% to 5%, aligned with practical missing-data guidance. Missingness will be summarised by centre, phase, outcome and covariate; compatibility with MCAR or MAR will be assessed using observed-data patterns. Multiple imputation by chained equations will be used for covariates with 5% to 40% missingness when appropriate, and covariates with more than 40% missingness will not be included in primary adjusted models. Location: Statistical analysis p.10, lines 4-10; reference 39.

13. The protocol would benefit from a clearer explanation of how adjustment variables were selected. A DAG or conceptual causal framework could strengthen the justification.

Response: We have added a conceptual framework description to Statistical analysis. It explains that covariates were selected a priori because patient characteristics, procedure characteristics, care context and calendar time may influence both monitoring outcomes and the distribution of cases across phases. The planned adjustment variables are now listed explicitly. Location: Statistical analysis p.9, lines 20-26; Table 3 p.18.

14. The discussion appropriately acknowledges limitations. However, residual confounding and inability to establish causality should be emphasized more strongly.

Response: We have strengthened the Discussion and the strengths/limitations box. The revised text states directly that the uncontrolled pre-post design cannot establish causality, that residual confounding may bias estimates, and that results will be interpreted as associations rather than causal effects. Location: Strengths box p.3, lines 33-34; Discussion p.11, lines 22-26.

15. Given the substantial financial and intellectual involvement of several authors, consider describing measures to ensure analytical independence and minimize investigator bias.

Response: We have added an 'Analytical independence and bias mitigation' subsection. It states that outcomes, thresholds and analyses were prespecified before comparative outcome analyses; statistical modelling will be implemented by statistics co-authors not involved in VPA invention or licensing; data extraction and quality-control steps will be documented; conflicts are fully disclosed; and Philips will have no role in data analysis, interpretation, manuscript preparation or publication decisions. Location: Analytical independence p.11, lines 1-6; Competing interests p.19, lines 32-36 and p.20, lines 1-7; Funding p.20, lines 8-17.

16. The term 'big-data study' in the title may be unnecessarily promotional.

Response: We have revised the title and terminology throughout, replacing 'big-data study' with 'large-scale observational study'. Location: Title p.1, lines 3-6; keywords p.2, lines 16-18.

17. The term 'sink-in phase' is uncommon. Consider replacing it with 'adaptation phase' or 'wash-in period'.

Response: We have replaced 'sink-in phase' with 'adaptation phase' throughout the manuscript. Location: Methods throughout, especially p.5, lines 3-12 and p.6, lines 1-9.

18. Please indicate whether the protocol follows relevant reporting guidelines for routinely collected health data studies (eg, RECORD extension of STROBE) and provide the checklist, if available.

Response: We have added a 'Reporting guidelines' subsection stating that the protocol follows TREND and RECORD where applicable. We provide a completed RECORD checklist as part of the revision package in addition to the TREND checklist requested by the Editorial Office. Location: Reporting guidelines p.5, lines 13-17; separate TREND/RECORD checklist file.

VERSION 2 - REVIEW

Reviewer	1
Name	Shahidi Delshad, Elham
Affiliation	Tehran University of Medical Sciences, Anesthesia
Date	20-Jul-2026
COI	

The authors have provided a thorough, comprehensive, and highly professional response to all my concerns.