

BMJ Open Evaluation of visual patient predictive for enhancing level 3 situation awareness: protocol for a multicentre randomised computer-based simulation and diagnostic accuracy study (true positive rate, precision, average lead time)

Cynthia Alexandra Hunn,¹ Hannes Bruns,¹ Sebastian Sahli,² Luca Wachtendorf,² Johannes Schäfer,³ Stefan Schwerin,⁴ Achilles Delis,⁵ Markus Kalisch,⁶ Gaspar Dugac,⁶ Arend Rahrisch,¹ Max Ebensperger,⁴ Arian Karimtar,² Gregor Massoth,⁵ Christopher Neuhaus,³ Alina Dubatovka,⁷ Christoph B Nöthiger,¹ G. Gasciauskaite,¹ Tadzio Raoul Roche,¹ David W Tscholl ¹

To cite: Hunn CA, Bruns H, Sahli S, *et al.* Evaluation of visual patient predictive for enhancing level 3 situation awareness: protocol for a multicentre randomised computer-based simulation and diagnostic accuracy study (true positive rate, precision, average lead time). *BMJ Open* 2025;**15**:e109171. doi:10.1136/bmjopen-2025-109171

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2025-109171>).

Received 10 August 2025
Accepted 13 December 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to

Dr David W Tscholl;
david.tscholl@usz.ch

ABSTRACT

Introduction Visual Patient Predictive (VPP) is an AI-based extension of the Visual Patient Avatar (VPA) that integrates deep learning models to predict upcoming vital sign deviations and display them as dashed visual elements. By explicitly showing anticipated changes, the system aims to support level 3 situation awareness—the projection of future patient states. This multicentre simulation study will evaluate whether predictive algorithms and visualisations integrated into the VPA (resulting in VPP) improve clinicians' ability to anticipate critical vital sign changes compared with conventional number-based and waveform-based monitoring and examine its effects on decision-making, confidence, workload and user acceptance.

Methods and analysis This investigator-initiated, randomised, within-subjects crossover, computer-based simulation trial will be conducted at five academic centres in Switzerland, Germany and the United States. Medical professionals from anaesthesiology departments will complete scenario-based prediction tasks using both VPP (as the index test) and conventional monitoring (as the reference standard) in randomised order, with the same participant evaluating both modalities and the identical underlying clinical scenario used in each condition, following video-based training and a learnability test. The primary outcome is recall (true positive rate) of vital sign deviation predictions. Secondary outcomes include average lead time, precision, prediction confidence, number and correctness of proposed interventions, perceived workload (NASA-TLX) and qualitative usability feedback. Quantitative data will be analysed using a logistic generalised linear mixed model with random intercepts for centre and participant, and a random slope for the intervention effect. Qualitative interviews will undergo thematic analysis.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Multicentre, randomised, computer-based simulation trial to evaluate a complete AI-supported predictive visualisation system for patient monitoring.
- ⇒ Within-subjects design controls for inter-individual variability and allows direct comparison between VPP and conventional monitoring.
- ⇒ Standardised video-based training and a preceding learnability test ensure consistent baseline knowledge across participants.
- ⇒ Combination of quantitative and qualitative methods provides a comprehensive evaluation of usability, decision-making impact and user acceptance.
- ⇒ Computer-based simulation in a controlled environment enables isolated assessment of the intervention but limits generalisability to real-world clinical settings.

Ethics and dissemination The leading ethics committee (Zurich, Switzerland; BASEC-Req-2023-00465) reviewed and approved the study protocol. Ethics committees at the other participating centres have obtained their respective approvals or waivers. Bonn: 2025-144-B0, Boston: 2025P000501, Heidelberg: S-376/2025, Munich: 2025-357 W-CB. As this simulation study involves only healthcare professionals performing prediction tasks based on simulated vital sign scenarios—without collection of patient data or any medically relevant personal data—it does not constitute human subjects research under applicable regulations. Study results will be disseminated through peer-reviewed publications and presentations at scientific conferences.

INTRODUCTION

Visual Patient Predictive (VPP) adds predictive algorithms and visualisations to the existing Visual Patient Avatar (VPA) concept, an avatar-based visualisation technology developed by the study authors' research group at the University Hospital Zurich, Switzerland, for patient monitoring that is commercially available as the *Philips Visual Patient Avatar* (Koninklijke Philips N.V., Amsterdam, The Netherlands).^{1–4} The Visual Patient concept translates numerical and waveform-based vital sign data into an intuitive, image-based representation and also displays the measurement status (eg, timed-out blood pressure readings). As the thresholds for visual indications are set more conservatively than those for auditory alarms, Visual Patient enables continuous visual monitoring of the patient's status, independent of auditory alarms. This permanent visual awareness supports the early detection of patient deterioration and ensures that patient and measurement status information remains continuously available, even when auditory alarms have been paused or deactivated.

The Visual Patient concept has been developed according to situation awareness-oriented design⁵ and validated through numerous preclinical, human-computer interaction studies,^{6–8} high-fidelity simulation studies^{9,10} and qualitatively in clinical practice.^{11–13} Results demonstrated faster and easier information acquisition, increased diagnostic confidence, reduced perceived cognitive workload and improved situation awareness, even under distraction,¹⁴ during the simultaneous monitoring of multiple patients,¹⁵ using peripheral vision¹⁶ and from greater distances.¹⁷ Qualitative user feedback suggests that the technology supports clinical workflows, particularly in the early recognition of multiple simultaneously deviating vital signs, the concurrent monitoring of multiple patients and communicating monitoring information to children and patients' relatives.

While highly effective for rapid situation awareness, the basic VPA is limited to depicting the current physiological state (eg, too high, too low or normal). It does not display trends or directional information, and upcoming deviations cannot be inferred from Visual Patient alone, as projections inherently require access to numerical time-series data. To address this limitation, and to provide predictive insights without reintroducing the cognitive burden associated with additional numerical displays for AI predictions, our group collaborated with the Swiss Federal Institute of Technology (ETH Zurich, Zurich, Switzerland) to develop deep learning models that analyse vital-sign streams in real time and forecast upcoming deviations. These predictions are integrated into the VPA using dashed predictive visualisations, which closely follow the established avatar visualisations. The resulting technology is VPP. Both the machine learning models and the predictive visualisations have been validated in separate studies and published in the *British Journal of Anaesthesia (BJA)* and *Diagnostics (Basel)*.^{18,19} An educational video illustrating the basic VPA and the complete

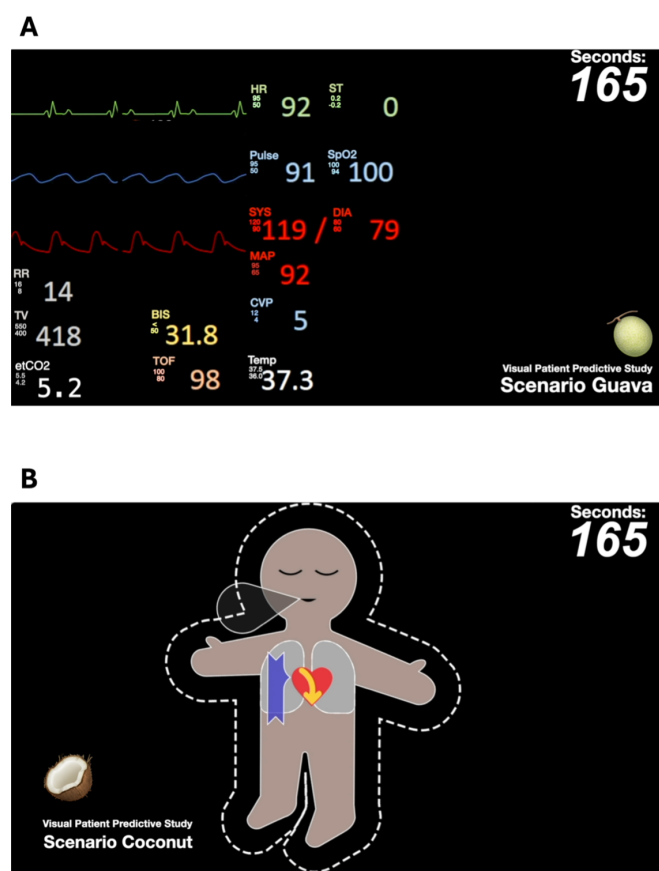


Figure 1 Example of display modalities used in the study. (A) Conventional monitoring format showing numeric values and waveform data for multiple vital signs (Scenario codename: Guava). (B) Corresponding scenario displayed using the VPP system, which presents vital sign information and prediction cues in an avatar-based format (Scenario codename: Coconut). Both images show time point 165 s in two different scenarios. To prevent participants from recognising repeated clinical cases across modalities, each scenario was assigned a unique fruit-based label, and no label was used more than once across visualisation formats. VPP, Visual Patient Predictive. BIS=Bispectral Index, TOF, train of four; CVP, Central venous pressure; MAP, Mean arterial pressure; SYS, Systolic; DIA, Diastolic; RR, Respiratory rate; TV, Tidal volume; etCO2, end-expiratory CO2 concentration.

AI-based VPP concept is available as online supplemental file 3. A representative example of a conventional and a VPP scenario used in this study is shown in figure 1.

Situation awareness, as classically defined by Mica Endsley, the founder of the concept, is divided into three levels: (1) perception of relevant information, (2) comprehension of its meaning, and (3) projection of future system states.^{5,20} These levels form the foundation for sound decision-making across all areas of human activity and are particularly critical in the clinical context. Breakdowns in situation awareness have been identified as the leading cause of human error, malpractice claims and cases reported in critical incident reporting

systems.^{21 22} Decisions drive performance: good decisions require good situation awareness, and good performance rests on good decisions. The VPA has already demonstrated improvements at levels 1 (perception) and 2 (comprehension), with users identifying more vital signs per unit of time and integrating them more quickly into correct diagnoses in computer-based and high-fidelity simulation studies.^{1 2 9}

Rationale and evidence gap

No previous study has investigated a fully integrated predictive visualisation system in which real-time AI-based forecasts are embedded directly into an established situation awareness-oriented monitoring interface. Existing research has separately validated the predictive algorithms¹⁸ and the visualisation concept,^{12 6–17} but not their combined impact on clinicians' ability to anticipate and respond to critical changes. Addressing this gap, the present study will be the first multicentre, randomised, computer-based simulation trial to evaluate whether the VPP system can improve Level 3 situation awareness and related decision-making compared with conventional monitoring.

METHODS AND ANALYSIS

Study Design

This protocol describes an investigator-initiated, multicentre, randomised, computer-based simulation study with a within-subjects crossover design, conducted at five study sites in Switzerland, Germany and the United States. The study aims to compare the accuracy of clinical personnel in predicting vital sign deviations based on clinical scenarios, which are presented either through conventional, number-based and waveform-based user interface resembling a standard state-of-the-art patient monitor without the basic VPA (constituting the reference standard), or the AI-based VPP (index test). A test of the learnability of VPP will also be performed. In addition, qualitative interviews will be conducted to assess acceptance, gather suggestions for improvement and explore potential future applications of the technology. See [figure 2](#) for an overview of the planned study proceedings.

Patient Public Involvement

Patients or members of the public were not involved in the design or conduct of this study. Given the patient-centred nature of the technology, we plan to involve patients in the presentation and dissemination of the research findings.

Participating Centres

The participating centres are the University Hospital Bonn (Germany), the Beth Israel Deaconess Medical Centre in Boston, MA (USA), the University Hospital Heidelberg (Germany), the University Clinic of the TUM (Germany), and the University Hospital Zurich (Switzerland).

Study Timeline

Planning and preparatory activities for the study commenced in the second quarter of 2023. The pilot study began at the Zurich centre in March 2025. The main study commenced in the second quarter of 2025, with data collection across all participating centres scheduled to be completed within 1 year. At the time of protocol submission, fewer than half of the planned participants had been enrolled, and no analyses for the main study had been performed. Statistical analyses will commence immediately after data collection is finalised, and submission of the first results manuscript is expected before the end of 2026.

Study Participants (Inclusion Criteria)

Participants in the study are medical professionals (residents, attending physicians and nurses) employed at the departments of anaesthesiology at the respective study centres and who regularly work with patient monitoring as part of their current clinical duties. An effort will be made to achieve a balanced distribution of the three professional groups within the overall cohort to adequately capture different perspectives and levels of experience in patient monitoring. The exact number of participants was determined through a sample size calculation based on the results of ten participants recruited at the Zurich site. These initial data collection served to estimate effect sizes and determine the sample size required for the overall study and will not be included in the main study analyses.

Study Setting

The study will be conducted in a quiet, controlled environment outside the participants' routine clinical duties. Participants will first watch a standardised, 10-min training video (online supplemental video 1), explaining the functionality of VPA and VPP. After completing the training video, participants can ask questions regarding the VPA, VPP visualisations, and the conventional monitor displays. These questions will be addressed immediately after the training, to ensure that all participants understand the presented material adequately before proceeding with the study.

All participants will subsequently complete a demographic survey and three consecutive phases of data collection (Learnability Test, Main Study, Interviews). This survey will collect, among other information, the name of the hospital, the participant's age, gender, professional role (resident, nurse or attending physician), years of professional experience and prior experience with the VPA platform.

Study Procedures and Outcomes

Learnability Test

Following the training, participants will complete a learnability test, representing a distinct study phase. They will be shown all individual visualisations of vital sign

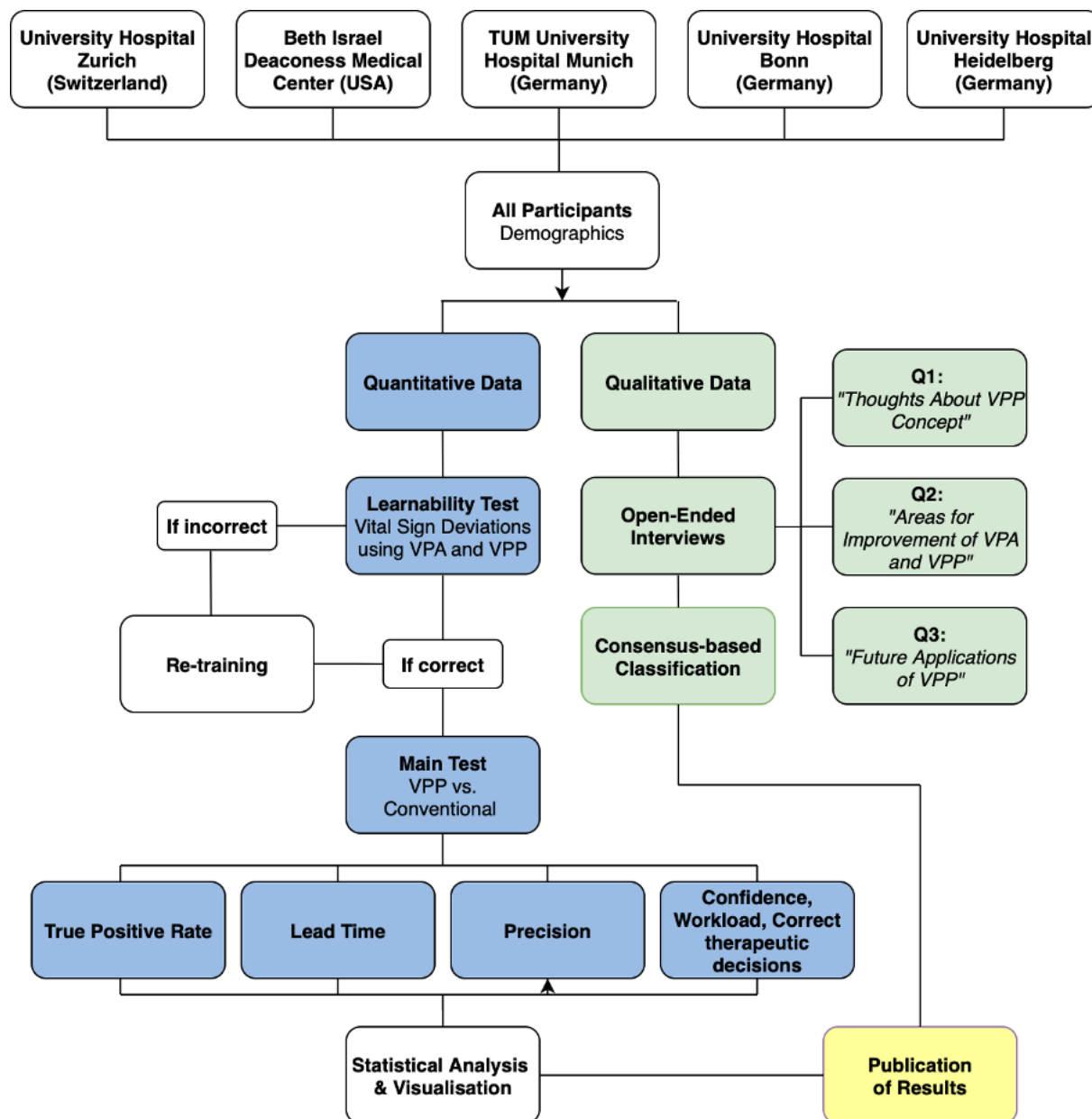


Figure 2 Study Flow Diagram showing quantitative and qualitative components of the multicentre evaluation of the VPP system. VPA, Visual Patient Avatar; VPP, Visual Patient Prediction.

deviations—both VPA and VPP—in randomised order and asked to identify the deviation depicted.

Participants will view the visualisations one by one and, after each visualisation, indicate which deviation they believe they have observed. If participants incorrectly identify a visualisation, the error will be recorded; however, participants will receive targeted retraining on the incorrectly identified visualisations before proceeding directly to the main study phase. The objective is to assess how accurately participants can interpret the visualisations after the standardised training.

Learnability Test Outcomes

The learnability test will assess secondary outcomes, as the study's primary outcome will be investigated in the main study conducted after this test. Specifically, in the

learnability test, the number of correctly identified visualisations will be measured, evaluated separately for each vital sign deviation ('too high' and 'too low') and aggregated across all visualisations. Additionally, the total number of correctly identified visualisations will be recorded separately for the non-predictive VPA and VPP. Results will be presented in a tabular or graphical form.

Rationale for Learnability Test Outcomes

This phase aims to determine how well participants can interpret VPA and VPP visualisations after a standardised training and to identify which visual elements are particularly well or poorly recognised. Findings will provide a systematic evaluation of the learnability of both technologies across a multicentre setting and inform targeted improvements of less intuitively interpretable

visualisations for potential future versions. The visualisations will not be modified before the main study, which takes place immediately afterwards.

Main Study

Data Collection Procedure

Participants will subsequently view four videos depicting clinical scenarios presented in randomised order. Each scenario is based on vital sign data constructed from real patient recordings, displayed once using VPP and once in conventional waveform-based and number-based form. To control for recognition and transfer effects, participants will not be informed that scenarios are repeated in different visualisation formats. Survey questions between scenarios serve as a cognitive distraction to minimise recall and learning effects.

Participants are seated individually at either a 13-inch or 15-inch Apple MacBook (Apple Inc., Cupertino, CA, USA) and view the scenarios together with a data collector. Their task is to predict upcoming vital sign deviations using both VPP and conventional monitoring. In the conventional scenarios, vital sign deviation thresholds are displayed next to the corresponding vital signs, allowing participants to base their predictions on clearly visible reference values.

Whenever a participant believes that a vital sign will deviate or recognises a prediction in the VPP condition, they press the space bar. The distinction reflects the two visualisation modalities: in the VPP condition, participants react to explicit predictive cues displayed by the avatar, whereas in the conventional condition, predictions are based solely on the participant's own clinical judgement without system support. In line with the Situation Awareness Global Assessment Technique (SAGAT), the scenario is then briefly paused to allow a standardised and unbiased assessment of situation awareness.⁵ During this pause, a brief structured form is completed together with the data collector using the iSurvey data collection tool (HarvestYourData.org, Wellington, New Zealand), a validated instrument frequently used by our research group in simulation-based studies.²³ The form records:

- ▶ Time of input (in seconds from the video start, based on the video timestamp),
- ▶ Vital sign expected to deviate,
- ▶ Direction of expected deviation (ie, 'increasing' or 'decreasing'),
- ▶ Subjective confidence in the prediction, rated from 0 (no confidence) to 100 (maximum confidence),
- ▶ Intended therapeutic intervention to prevent the deviation (selected from the list provided in online supplemental appendix 1).

After data entry, the data collector clicks 'next' in the electronic survey, and the scenario continues until the participant makes another input. If a participant makes multiple consecutive predictions concerning the same vital sign within a scenario, without an intervening deviation, only the first entry will be considered for analysis. This approach captures the participant's initial situational

impression and emphasises their immediate recognition and projection capability. As the study aims to assess Level 3 situation awareness and the ability to make appropriate hypothetical clinical decisions—prioritising the first prediction reflects the most authentic and intuitive judgement. If a participant makes a prediction in the opposite direction for the same vital sign, only the last stated direction will be considered the final prediction. The occurrence of a deviation in a vital sign will invalidate all previously made predictions concerning that parameter.

Scenario Description

The scenarios were created using vital sign data from the publicly available *vitaldb.org* database²⁴ to simulate typical clinical decision-making situations. To minimise situational priming through coherent clinical patterns, vital signs from different patients were used across the scenarios. This approach allows the effect of the visualisations to be assessed independently of participants' pattern matching, while ensuring that the machine learning models are tested on realistic clinical data.

Clinical case vignettes were drafted to demonstrate that the situations depicted could plausibly occur in clinical practice (online supplemental appendix 1). Subsequently, the machine learning models were applied to the completed scenarios. The scenarios were provided once for the application of the machine learning models and were not modified to optimise model performance. A tabular overview of the vital sign changes and predictions shown in the scenarios is available in online supplemental appendix 1.

Primary and Secondary Outcomes (Prediction Scoring and Evaluation Metrics)

- ▶ For each participant and scenario, we evaluate predictions in the following way:
- ▶ Vital signs are handled independently, and results are aggregated.
- ▶ Prediction blocks are defined in an event-specific manner: consecutive predictions indicating the same direction of vital sign deviation (ie, 'increasing' or 'decreasing') are grouped into prediction blocks, continuing up to (but not across) the next deviation event.
- ▶ The timestamp of the first prediction in each block (the 'block start') marks the moment when the participant committed to that directional assessment. For each event, the active block is the last prediction block (if any) that ends before that event. An active block is counted as true if its direction matches the upcoming event; otherwise, it is counted as false.
- ▶ The timestamp of the first prediction in each block (the 'block start') marks the moment when the participant committed to that directional assessment. For each event, the active block is the last prediction block (if any) that ends before the event occurs. An active block is counted as true if its direction matches the upcoming event; otherwise, it is counted as false.

- For the primary outcome, it does not matter at which exact moment within an active block the prediction is made, if the prediction occurs during that block.
- For the secondary outcome (average lead time), however, the exact timing within the block is relevant, as lead time is calculated based on the block start relative to the onset of the deviation event.

Primary outcome

- True positive rate (TPR; also called ‘recall’).

Let (T) denote the number of true active blocks over all vital signs and let (N_e) be the total number of actual events across all vital signs.

Then we define: $TPR = \frac{T}{N_e}$

Secondary outcomes

- Average Lead-Time (ALT)

If there are any true active blocks, we calculate the time difference between the start of each such block and the corresponding event, resulting in an individual lead-time (L_i)

$$L_i = event_i - block_start_i, \text{ for } i = 1, \dots, n$$

The (ALT) is then computed as:

$$ALT = \frac{1}{n} (L_1 + L_2 + \dots + L_n)$$

- Precision (P)

Let (T) be the number of true active blocks over all vital signs (as for TPR) and (N_p) be the number of all prediction blocks over all vital signs. Then we define:

$$P = \frac{T}{N_p}$$

As N_p includes all prediction blocks—both correct and incorrect—false predictions directly reduce the precision score. This metric therefore reflects the proportion of correct predictions among all predictions made.

Further Secondary Outcomes

- Participants’ subjective confidence in their predictions, measured using a continuous confidence rating scale ranging from 0 to 100%
- Number of correctly selected therapeutic interventions per scenario, compared between conditions, accounting for differences in the total number of decisions.
- Total number of therapeutic interventions selected per scenario, irrespective of correctness, to assess decision-making behaviour with and without VPP.
- Perceived workload per scenario, assessed using the NASA-TLX.

Subjective workload will be assessed using the validated NASA Task Load Index (NASA-TLX), an internationally recognised instrument for measuring perceived cognitive workload. Our research group has specifically validated the NASA-TLX for tasks involving patient monitoring, ensuring context-specific applicability.²⁵

Rationales for Outcomes

The choice of outcomes is based on Mica Endsley’s model of situation awareness. The primary aim is to evaluate the effect of VPP on situation awareness Level 3 (projection) and clinical decision-making.

The primary outcome—the number of correctly predicted critical deviations—will determine whether using VPP leads to a higher number of correct predictions. The time interval between prediction and actual deviation will help assess whether VPP enables earlier, and therefore potentially more clinically relevant, predictions.

The number of correctly selected therapeutic interventions aims to demonstrate the link between improved situation awareness and better clinical decision-making. Participants’ subjective confidence in their predictions will also be measured, as higher confidence could enhance action readiness and reduce decision-making delays. Finally, perceived mental workload will be assessed, as elevated workload is a known risk factor for diminished situation awareness.⁵

Qualitative Interviews

After completing the scenario study phase, participants will take part in a qualitative individual interview to gain qualitative feedback about the studied technologies. The interviews are open-ended and guided by the following three core questions:

1. “Please share freely your thoughts about the Visual Patient Predictive concept.”
2. “How could Visual Patient Predictive be used in the future?”
3. “How could Visual Patient and Visual Patient Predictive be improved?” (eg, visualisation, concept and new ideas)

The interviews aim to gather information on acceptance, suggestions for improvement and potential future applications.

The qualitative responses will be analysed using a thematic analysis approach.²⁶ If participants’ answers are not recorded in English, they will be translated sentence by sentence into English using ChatGPT (OpenAI, San Francisco, USA - the latest available model at the time of translation). Human authors will then verify the accuracy of these machine translations. The answers will be categorised into topics and analysed accordingly in the next step. The authors have extensive experience with this methodological approach.

Responses to the qualitative interview questions will be transcribed in parallel by the study investigators on a MacBook Air using *Microsoft Word* (Microsoft Corporation, Redmond, WA, USA) during the interviews.

Rationales for Interviews

Qualitative data collection is undertaken because qualitative feedback is an essential component of successful product development. The qualitative findings can be integrated with the quantitative data and interpreted jointly, enabling targeted incorporation of insights from user experience into future versions of VPA and VPP in accordance with the principles of user-centred design.

Data Analysis and Statistical Considerations

Before the main study began, a pilot study was conducted at the Zurich site, which served as the basis for the subsequent sample size calculation.

Statistical Model for primary endpoint

We analysed the primary binary end-point (correct prediction=yes/no for each time-window) with a logistic generalised linear mixed model (GLMM). The model contains the following:

- ▶ a fixed effect for avatar (yes vs no) – the parameter of interest;
- ▶ fixed effects for scenario (A vs B) and period, ie, the order in which the combination of avatar and scenario appeared (1–4);
- ▶ a random intercept and a random avatar-specific slope for each doctor (to allow doctors to differ both in baseline recall and in how much they benefit from the avatar);
- ▶ a random intercept for centre (to accommodate between-centre heterogeneity).

The GLMM is fitted with `lme4::glmer()` and the avatar effect is tested with a two-sided Wald z-test at $\alpha=0.05$.

Sample Size Estimation and Power analysis

We powered the study to detect $\beta(\text{avatar})=\log(1.25)$, that is, a 25% increase in the odds of a correct prediction of an event when the avatar is shown (OR 1.25) with 80% power. Using parameter values derived from the pilot study (see online supplemental appendix 2 for details), we ran a Monte-Carlo simulation with 500 replicates for several care provider counts. For 5 centres $\times$ 10 care providers (50 care providers total) the GLMM analysis achieved power=0.83 (95 % Monte-Carlo CI 0.80 to 0.87), exceeding the 80% target. Therefore 10 care providers per centre was chosen for the trial.

Further Statistical Procedure

Multivariate analyses will be conducted to investigate the influence of participants' demographic characteristics on the learnability and recognition of the visualisations, their ability to make accurate predictions and decision-making performance with both technologies. Factors to be examined include prior monitoring experience, professional role, previous exposure to the VPA and whether the VPA is used at the participant's institution.

Outcome variables will include binary outcomes (eg, correct vs incorrect predictions defined by correct vital sign and direction), continuous measures (eg, time to correct prediction), and ordinal ratings (eg, confidence levels). Where appropriate, generalised linear models or mixed-effects models will be applied to account for repeated measures and within-subject dependencies.

Full pilot study results and statistical code are available in online supplemental appendix 2.

Qualitative Interviews

Responses from qualitative interviews will be analysed using thematic analysis. Two independent coders will

identify themes, which will be discussed and refined in an iterative process.

Ethics and Dissemination

The responsible ethics committees at the participating centres reviewed the study protocol and issued declarations of non-responsibility or waivers. Bonn: 2025–144-BO, Boston: 2025P000501, Heidelberg: S-376/2025, Munich: 2025–357 W-CB, Zurich: BASEC-Req-2023–00465.

As this simulation study involves only healthcare professionals performing prediction tasks based on simulated vital sign scenarios—without the collection of patient data or any other medically relevant personal data—it does not constitute human subjects research under applicable regulations and poses no risk to patients. All participants will provide written informed consent before participation. Data will be collected and stored in accordance with institutional policies and applicable data protection regulations, and no individual participant will be identifiable in any publication.

The study results will be disseminated through peer-reviewed journal articles and presentations at relevant scientific conferences. Separate publications are planned for the quantitative results and for the qualitative findings.

Data Deposition and Curation

De-identified quantitative datasets and associated statistical code, representing the main part of the study, will be securely stored at the coordinating centre. These materials will be made available on reasonable request to the corresponding author, subject to institutional policies and applicable data protection regulations. Anonymised excerpts from the qualitative interviews will be included in the qualitative results publication, and full anonymised interview transcripts will likewise be stored securely and made available on request, in accordance with participant consent and data protection requirements.

Strengths and Limitations

This study presents several inherent strengths and limitations. Strengths include the clear operationalisation of outcome measures, the use of a standardised training video with a preceding learnability test, the direct comparison with conventional monitoring and the targeted focus on Level 3 situation awareness—a relatively underexplored aspect of human–machine interaction in medicine. Including five international study centres provides a certain degree of generalisability, and the combination of quantitative and qualitative methods allows for a comprehensive evaluation of the technology from both clinical and user-centred perspectives.

At the same time, important limitations exist. First, this is a computer-based simulation study conducted in a controlled, distraction-free environment. While this setting enables an isolated assessment of the VPP technology without the influence of external clinical factors or cognitive overload, it does not fully reflect the complexity and dynamics of real-world clinical practice.

Furthermore, the assessment relies on brief scenario pauses in accordance with the Situation Awareness Global Assessment Technique (SAGAT).⁵ Although widely used in simulation research, such pauses do not occur in clinical workflows and may only approximate the dynamic decision-making processes of practising clinicians. This study therefore represents a foundational first step, which—if feasible—should be followed by future research evaluating the system's performance using real patient data in high-fidelity simulations and ultimately evaluate patient outcomes in actual clinical settings.

Second, the scenarios tested are limited to only two case examples constructed from real vital sign data from different patients. These are presented without a broader clinical context. While this reduces the influence of clinical pattern recognition and allows for a focused analysis of the visualisations, it limits case variability and the realism of the decision-making process.

Third, the study is conducted exclusively with anaesthesia personnel at five academic centres in Switzerland, Germany and the United States. While this provides a solid foundation for initial insights into generalisability, it excludes other specialties, levels of experience and technological backgrounds.

In addition, this study does not assess the technical performance or calibration of the underlying predictive models. A separate validation study—building on the initial validation already performed—will be required to evaluate model performance across large, heterogeneous datasets and varying threshold settings.

Further limitations include the absence of long-term learning evaluation and potential changes in usage quality over time, as participants are tested immediately after their first exposure to the technology. Additionally, there may be a self-selection bias, as particularly motivated or technologically inclined individuals may be more likely to participate.

Despite these limitations, this study represents an important first step in investigating AI-supported predictive visualisation technologies, with the potential to lay the groundwork for future research in high-fidelity simulation and clinical application.

Author affiliations

¹Institute for Anesthesiology and Perioperative Medicine, University Hospital Zurich, Zurich, Switzerland

²Department of Anesthesia, Critical Care and Pain Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

³Department of Anaesthesiology, Heidelberg University, Heidelberg, BW, Germany

⁴Department of Anesthesiology and Intensive Care, TUM, Munich, BY, Germany

⁵Department of Anesthesiology, Universitätsklinikum Bonn, Bonn, NRW, Germany

⁶Seminar for Statistics, ETH Zurich, Zurich, ZH, Switzerland

⁷Department of Computer Science, ETH Zurich, Zurich, ZH, Switzerland

Contributors CAH: Conceptualisation, Methodology, Software; HB: Conceptualisation, Methodology, Validation, Writing – Original Draft, Writing – Review and Editing, Formal analysis, Data curation; SSA: Investigation, Writing – Review and Editing; LW: Investigation; JS: Investigation; SSC: Investigation, Resources; ADE: Investigation, Resources, Writing – Review and Editing; MK: Conceptualisation, Methodology, Validation, Formal analysis, Supervision; GD: Conceptualisation, Methodology, Validation, Formal analysis; AR: Investigation,

Writing – Review and Editing; ME: Investigation, Writing – Review and Editing, Visualisation; AK: Investigation, Supervision; GM: Investigation, Resources, Writing – Review and Editing, Supervision; CN: Investigation, Resources, Supervision; ADU: Conceptualisation, Software; CBN: Conceptualisation, Investigation; GG: Conceptualisation, Software; TRR: Conceptualisation, Methodology, Validation, Writing – Review and Editing; DWT: Conceptualisation, Methodology, Software, Validation, Resources, Writing – Original Draft, Writing – Review and Editing, Visualisation, Supervision, Project administration, Funding acquisition; DWT is the guarantor. Generative AI (ChatGPT, OpenAI, San Francisco, CA, USA) was used to enhance the clarity and readability of the manuscript. All AI-assisted revisions were reviewed and approved by the authors to ensure accuracy and completeness.

Competing interests CBN, TRR and DWT are inventors of the Visual Patient Predictive technology. This technology has been licensed by the University of Zurich to Philips. In accordance with University of Zurich regulations, inventors are entitled to a share of potential future revenues in the event of successful commercialization. CBN and DWT are also inventors of the Visual Patient Avatar technology, on which Visual Patient Predictive is based, and receive royalties related to this technology. Authors CAH, CBN, TRR and DWT have received honoraria for lectures from Philips. Authors CAH, AD, TRR and DWT have received research funding from Philips through joint development agreements with their respective universities. All funding and remuneration were managed via institutional agreements, and Philips had no role in the study design, data collection, data analysis, data interpretation, or manuscript preparation.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the leading ethics committee (Zurich, Switzerland; BASEC-Req-2023-00465) reviewed and approved the study protocol. Ethics committees at the other participating centres have obtained their respective approvals or waivers. Bonn: 2025-144-B0, Boston: 2025P000501, Heidelberg: S-376/2025, Munich: 2025-357-W-CB. As this simulation study involves only healthcare professionals performing prediction tasks based on simulated vital sign scenarios—without the collection of patient data or any other medically relevant personal data—it does not constitute human subjects research under applicable regulations and poses no risk to patients. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. The results of the pilot study are included in the supplementary material. Beyond this, no data have been curated or analyses performed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

David W Tscholl <https://orcid.org/0000-0002-3677-3705>

REFERENCES

- 1 Tscholl DW, Handschin L, Neubauer P, *et al*. Using an animated patient avatar to improve perception of vital sign information by anaesthesia professionals. *Br J Anaesth* 2018;121:662–71.
- 2 Tscholl DW, Rössler J, Said S, *et al*. Situation Awareness-Oriented Patient Monitoring with Visual Patient Technology: A Qualitative Review of the Primary Research. *Sensors (Basel)* 2020;20:2112.

- 3 Philips. Philips' visual patient avatar pioneers 'at-a-glance' patient data for faster, better decision-making in the operating room: philips. 2023. Available: <https://www.philips.com/a-w/about/news/archive/standard/news/press/2023/20231013-philips-visual-patient-avatar-pioneers-at-a-glance-patient-data-for-faster-better-decision-making-in-the-operating-room.html>
- 4 Gascauskaitė G, Lunkiewicz J, Roche TR, *et al*. Human-centered visualization technologies for patient monitoring are the future: a narrative review. *Crit Care* 2023;27:254.
- 5 Endsley MRJ. *Designing for situation awareness: An approach to user-centered design*. 3rd edn. Boca Raton, FL, USA: CRC Press, Inc, 2025.
- 6 Bergauer L, Braun J, Roche TR, *et al*. Avatar-based patient monitoring improves information transfer, diagnostic confidence and reduces perceived workload in intensive care units: computer-based, multicentre comparison study. *Sci Rep* 2023;13:5908.
- 7 Tscholl DW, Rössler J, Handschin L, *et al*. The Mechanisms Responsible for Improved Information Transfer in Avatar-Based Patient Monitoring: Multicenter Comparative Eye-Tracking Study. *J Med Internet Res* 2020;22:e15070.
- 8 Wetli DJ, Bergauer L, Nöthiger CB, *et al*. Improving Visual-Patient-Avatar Design Prior to Its Clinical Release: A Mixed Qualitative and Quantitative Study. *Diagnostics (Basel)* 2022;12:555.
- 9 Roche TR, Said S, Braun J, *et al*. Avatar-based patient monitoring in critical anaesthesia events: a randomised high-fidelity simulation study. *Br J Anaesth* 2021;126:1046–54.
- 10 Akbas S, Said S, Roche TR, *et al*. User Perceptions of Different Vital Signs Monitor Modalities During High-Fidelity Simulation: Semiquantitative Analysis. *JMIR Hum Factors* 2022;9:e34677.
- 11 Hunn CA, Lunkiewicz J, Nöthiger CB, *et al*. Qualitative Exploration of Anesthesia Providers' Perceptions Regarding Philips Visual Patient Avatar in Clinical Practice. *Bioengineering (Basel)* 2024;11:323.
- 12 Lunkiewicz J, Fries D, Milovanovic P, *et al*. Pediatric Anesthesia Providers' Perspective on the Real-Life Implementation of the Philips Visual Patient Avatar: A Qualitative Study. *Children (Basel)* 2023;10:1841.
- 13 Radaviciute I, Hunn CA, Lunkiewicz J, *et al*. Survey-based qualitative exploration of user perspectives on the philips visual patient avatar in clinical situation management. *Sci Rep* 2024;14:22176.
- 14 Pfarr J, Ganter MT, Spahn DR, *et al*. Effects of a standardized distraction on caregivers' perceptive performance with avatar-based and conventional patient monitoring: a multicenter comparative study. *J Clin Monit Comput* 2020;34:1369–78.
- 15 Garot O, Rössler J, Pfarr J, *et al*. Avatar-based versus conventional vital sign display in a central monitor for monitoring multiple patients: a multicenter computer-based laboratory study. *BMC Med Inform Decis Mak* 2020;20:26.
- 16 Pfarr J, Ganter MT, Spahn DR, *et al*. Avatar-Based Patient Monitoring With Peripheral Vision: A Multicenter Comparative Eye-Tracking Study. *J Med Internet Res* 2019;21:e13041.
- 17 Milovanovic P, Braun J, Hunn CA, *et al*. Avatar-based versus conventional patient monitoring with distant vision: a computer-based simulation study. *J Clin Monit Comput* 2025;39:1065–75.
- 18 Dubatovka A, Nöthiger CB, Spahn DR, *et al*. Predicting vital sign deviations during surgery from patient monitoring data: developing and validating single-stream deep learning models. *Br J Anaesth* 2024;133:889–92.
- 19 Malorgio A, Henckert D, Schweiger G, *et al*. Using Visual Patient to Show Vital Sign Predictions, a Computer-Based Mixed Quantitative and Qualitative Simulation Study. *Diagnostics (Basel)* 2023;13:3281.
- 20 Schulz CM, Endsley MR, Kochs EF, *et al*. Situation awareness in anesthesia: concept and research. *Anesthesiology* 2013;118:729–42.
- 21 Schulz CM, Burden A, Posner KL, *et al*. Frequency and Type of Situational Awareness Errors Contributing to Death and Brain Damage: A Closed Claims Analysis. *Anesthesiology* 2017;127:326–37.
- 22 Schulz CM, Krautheim V, Hackemann A, *et al*. Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. *BMC Anesthesiol* 2016;16:4.
- 23 Tscholl DW, Weiss M, Spahn DR, *et al*. How to Conduct Multimethod Field Studies in the Operating Room: The iPad Combined With a Survey App as a Valid and Reliable Data Collection Tool. *JMIR Res Protoc* 2016;5:e4.
- 24 Lee HC, Park Y, Yoon SB, *et al*. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. *Sci Data* 2022;9:279.
- 25 Said S, Gozdzik M, Roche TR, *et al*. Validation of the Raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: Pooled Analysis Study Using Mixed Models. *J Med Internet Res* 2020;22:e19472.
- 26 Braun J, Clarke V, Hayfield N, *et al*. Thematic analysis. In: Liamputtong P, ed. *Handbook of research methods in health social sciences*. Singapore: Springer Singapore, 2019: 843–60.