

Review Article

Autonomous Operating Room Monitoring Platforms: A Multicenter Evaluation

Anna Nagy^{1*}, Karolien Brenner² and Hanna Colombo³

¹Department of Critical Care Medicine, Alpine Institute of Medical Sciences, Austria

²Department of Anesthesiology, Tyrrhenian University of Health Sciences, Italy

³Division of Anesthesia and Intensive Care, Tyrrhenian University of Health Sciences, Italy

ABSTRACT

Ceiling-mounted cameras, ambient microphones, and depth sensors are increasingly deployed in operating rooms as part of platforms marketed as autonomous monitoring systems, capable of tracking surgical phases, safety-checklist compliance, team traffic, and noise in real time.

This review surveys the multicenter evidence for these platforms, tracing their conceptual lineage from aviation-style 'black box' recorders and decades-old automated anesthesia record keeping to current computer-vision and audio-analytic systems.

Observational multicenter data show a consistent gap between assumed and actual safety-behavior compliance once continuous machine observation replaces self-report, along with measurable team traffic and distraction patterns not previously quantifiable at scale.

The alarm-fatigue literature from bedside physiologic monitoring offers a direct cautionary parallel, and the platforms with the most credible evidence to date favor aggregated feedback over indiscriminate real-time alerting.

We conclude that current evidence supports autonomous monitoring as a measurement and quality-improvement instrument rather than a validated driver of improved hard clinical outcomes, and we outline the generalizability, privacy, and alarm-management questions that remain open.

Keywords: operating room monitoring; surgical safety checklist; artificial intelligence; alarm fatigue; perioperative quality improvement

Citation: Nagy A, Brenner K and Colombo H (2026) Autonomous Operating Room Monitoring Platforms: A Multicenter Evaluation. *Anesthesiology*. 6(1):01-05.

DOI: 10.17140/2026.vol6.i01.pp01-05

Corresponding Author: Anna Nagy

E-mail Address: a.nagy@aims.at

Article History: Received 08 April 2026 | Accepted 07 May 2026 | Published 03 June 2026

Copyright: © 2026 Anna Nagy, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<https://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Walk into almost any modern operating room and the instrumentation looks, at first glance, much as it did fifteen years ago: a physiologic monitor, an anesthesia

workstation, a surgical field, a scrub team moving with practiced choreography. What has changed is largely invisible to the untrained eye. Ceiling-mounted cameras, ambient microphones, door and traffic sensors, and depth-sensing units now quietly record

the room's activity in many academic centers, feeding video, audio, and kinematic data into platforms that were unimaginable a decade ago [1]. These systems are increasingly described, sometimes with more marketing enthusiasm than precision, as "autonomous" operating room monitors - platforms capable of recognizing surgical phases, flagging safety-checklist deviations, counting instruments, detecting door openings during critical portions of a case, and in some early implementations, alerting staff to hazardous events in near real time [2].

The impulse behind this technology is not new. Anesthesiologists have used automated record-keeping systems since the late 1980s specifically because manual charting under load is unreliable and prone to omission [3]. What is new is the scope of what can now be captured and analyzed - not just vital signs, but the entire sociotechnical environment of the room: who is present, what is being said, how instruments move, and whether protocol steps are actually performed as written rather than merely charted as performed. This review surveys the current state of autonomous and semi-autonomous OR monitoring platforms, the multicenter evidence supporting (or complicating) their adoption, and the practical, ethical, and human-factors questions that anesthesiologists and perioperative leaders are now confronting as these systems move from pilot units into routine service.

FROM BLACK BOXES TO LEARNING SYSTEMS

The aviation-inspired "black box" concept is probably the most visible entry point for this technology into clinical practice. Modeled explicitly on flight data recorders, OR black box platforms combine synchronized audio, video, and device data streams to reconstruct a case in full after the fact, and increasingly to surface deviations during the case itself [4]. Early deployments concentrated on retrospective review - a surgical team or quality committee replaying a case to understand a complication or a near miss - but the more consequential shift has been toward continuous, prospective auditing across many rooms and many sites simultaneously, something no human observer could sustain [5].

A multicenter analysis using this kind of platform to remotely audit real-world adherence to the surgical safety checklist illustrates both the promise and the discomfort of the approach: automated review found substantially lower rates of complete, verbally confirmed checklist performance than institutions had assumed from self-report, even at sites that considered checklist culture mature [6]. That gap between believed and observed compliance is, in a sense, the entire argument for these platforms - human self-assessment of routine safety behavior is optimistic, and continuous machine observation is not. Professional society coverage of these deployments

has emphasized similar findings across multiple surgical specialties and multiple hospital systems, with case-level analytics used to identify recurring workflow inefficiencies (turnover delays, instrument tray mismatches, non-therapeutic OR traffic) as well as safety-relevant events [7].

The technical anatomy of a modern platform typically layers several distinct capabilities. Computer vision models trained on surgical video can segment a procedure into phases, recognize instruments in the field, and in some implementations estimate technical skill or detect atypical movement patterns associated with adverse events [8]. Separate models applied to ambient audio can flag elevated noise, cross-talk during critical phases, or the number of non-case-relevant conversations - each associated in prior human-factors literature with distraction-related error, though causally difficult to isolate [9]. Depth-sensing cameras, which capture room geometry and body position without recording identifiable video, have been proposed specifically to address the privacy objections that have slowed adoption of full-video systems, with automated workflow-recognition accuracy from depth data alone now approaching that of conventional video pipelines in early comparative work [10].

TEAM BEHAVIOR, SITUATIONAL AWARENESS, AND TRAFFIC

Beyond checklist compliance, a growing body of work has used computer vision to model something considerably harder to define: the situational awareness of the surgical team as a group. Rather than treating the OR as a single patient-monitor dyad, these systems track the position, gaze direction (where instrumented), and movement of every person in the room to infer whether the team is collectively attending to the field, to the anesthesia workstation, or to unrelated activity, and whether attention degrades at predictable points in a case such as handoffs or unplanned interruptions [11]. This line of work sits adjacent to a long tradition of manual observational studies of OR traffic and door openings, which have consistently linked excess traffic to both infection risk and workflow disruption, but which were previously limited to small samples collected by human observers over a handful of shifts. Continuous sensor-based traffic counting removes that sampling limitation and allows traffic to be correlated, case by case, with case duration, case complexity, and downstream outcomes rather than treated as a generic quality metric.

The distraction and noise literature that these systems now formalize predates the technology considerably. Professional guidance from anesthesiology societies has long recognized that non-clinical conversation, personal electronic device use, and elevated ambient noise during critical phases of anesthetic care - induction, emergence, and any period of hemodynamic instability - carry real risk, and has

called for institutional policies limiting non-essential activity during those windows [12]. What autonomous monitoring adds is not a new hazard model but a new measurement instrument: rather than relying on incident reports or observer studies, hospitals can now quantify, across hundreds of cases, exactly how often critical-phase distraction actually occurs, in which rooms, with which teams, and whether an intervention (a "sterile cockpit" policy, a redesigned handoff script) measurably changes the pattern rather than merely being adopted on paper.

ALARM BURDEN AND THE LIMITS OF AUTOMATED VIGILANCE

Any discussion of autonomous OR monitoring has to reckon honestly with alarm fatigue, because it is the clearest cautionary tale for what happens when automated vigilance outpaces human capacity to act on it. Physiologic monitors alone generate alarm rates in some ORs and ICUs high enough that the overwhelming majority are acknowledged reflexively or silenced rather than individually assessed, a pattern documented across more than a decade of biomedical instrumentation literature [13]. The Anesthesia Patient Safety Foundation has repeatedly flagged alarm fatigue as a contributor to missed critical events specifically because desensitization is a predictable consequence of high false-alarm rates, not a failure of individual vigilance [14]. Professional society guidance on alarm management for anesthesia professionals has pushed institutions toward alarm customization, tiered alerting by clinical severity, and default settings validated against actual event rates rather than manufacturer defaults calibrated for the most conservative possible threshold [15].

The lesson transfers directly to room-level autonomous monitoring. A platform that generates a checklist-deviation alert, a noise alert, and a traffic alert on every case, indiscriminately, will be tuned out by staff within weeks regardless of how sophisticated its underlying models are - the human-factors failure mode is identical whether the signal originates from a pulse oximeter or a ceiling camera. The more credible implementations reported to date have therefore emphasized asynchronous, aggregated feedback (weekly or monthly team-level dashboards) over real-time alerting for lower-acuity findings, reserving synchronous alerts for a narrow set of high-specificity events such as a retained-instrument risk signal or a hard door-opening violation during a documented critical phase [16]. Whether this restraint holds as vendors compete on feature breadth, or whether OR monitoring simply reproduces the alarm-fatigue trajectory of bedside physiologic monitoring at the room level, is one of the more consequential open questions in this literature.

ACCURACY, GENERALIZABILITY, AND THE MULTICENTER EVIDENCE PROBLEM

Model performance figures reported from single-institution pilot studies do not reliably predict performance elsewhere, and this is arguably the central methodological weakness of the current evidence base. Surgical phase-recognition and instrument-detection models trained on video from one health system's equipment, room lighting, camera placement, and case mix have shown measurable accuracy drops when deployed at a second site without retraining, a pattern documented across several systematic reviews of AI models applied to OR video [2]. This is not a minor technical footnote; it is the difference between a platform that can be marketed nationally and one that requires substantial site-specific calibration before its outputs can be trusted for anything beyond retrospective, human-reviewed case analysis.

Multicenter evaluations that do exist have generally been observational rather than randomized, comparing checklist compliance, turnover time, or traffic metrics before and after platform deployment across a handful of hospitals within the same health system or vendor partnership [6]. This design is defensible given the practical and ethical difficulty of randomizing an OR-wide surveillance intervention, but it leaves open substantial confounding - deployment of a monitoring platform is rarely an isolated intervention; it typically accompanies renewed leadership attention to OR safety culture, refreshed checklist training, and Hawthorne-type effects from staff awareness of being observed, any of which could independently move the outcomes attributed to the software. Broader reviews of AI and machine learning applied to patient safety in complex surgery have made a similar point: the evidence is accumulating quickly in volume but is still concentrated in feasibility and single-arm before/after designs, with few sites reporting on hard clinical outcomes such as complication rates or mortality rather than process measures [17].

Legal, regulatory, and ethical questions have moved somewhat more slowly than the technology itself. Recording identifiable audio and video of clinical staff and, depending on jurisdiction and consent framework, of patients under anesthesia raises questions about discoverability in litigation, staff consent and labor-relations concerns, and data governance that differ meaningfully across institutions and countries, and that vendors and hospital counsel are still working through case by case [18]. Depth-sensing and other de-identified sensing modalities have been proposed partly as a technical answer to this problem, trading some analytic richness for a materially smaller privacy footprint, and early comparative accuracy data suggest that trade-off may be smaller than initially feared for at least some workflow-recognition tasks [10].

WHERE THE TECHNOLOGY IS HEADED

The trajectory across recent literature and trade press

coverage points toward integration rather than replacement - autonomous monitoring is being positioned as a layer that sits alongside existing anesthesia and surgical workflows, feeding structured data into quality-improvement programs and, in some vendor roadmaps, into intraoperative decision support rather than functioning as a standalone alerting product [19]. Broader reviews of AI and robotics in general surgery describe a similar convergence: instrument tracking, phase recognition, and outcome prediction models are increasingly bundled with surgical robotics platforms rather than sold as independent monitoring add-ons, which may accelerate adoption but also concentrates data governance and liability questions with a smaller number of vendors [20]. Preliminary work on real-time quality-assurance systems that flag protocol deviations as they occur, rather than in retrospective review, suggests the next phase of development is aimed squarely at synchronous intervention - catching a missed step while it can still be corrected rather than documenting it afterward for a morbidity and mortality conference [21].

For anesthesiologists specifically, the more immediate relevance may be less about the headline safety-checklist and instrument-tracking use cases and more about what continuous room-level sensing can reveal about the anesthesia workspace itself: workstation alarm burden correlated against actual room activity, distraction windows mapped against induction and emergence timing, and traffic and noise data that can finally test, with real numbers rather than observational studies, whether sterile-cockpit-style policies during critical phases of anesthetic care actually change behavior when they are adopted [12]. That capacity to measure what was previously only inferred is, more than any single alerting feature, what distinguishes this generation of OR monitoring platforms from the automated record-keeping systems anesthesiology has used for decades [3].

CONCLUSION

Autonomous operating room monitoring has moved, over a relatively short span, from a research curiosity into a real if still unevenly distributed feature of perioperative practice. The multicenter evidence assembled so far supports a fairly specific and modest claim: continuous, sensor-based observation reveals real gaps between assumed and actual safety-behavior compliance, and it can quantify team dynamics, traffic, and noise in ways manual audit never could. It does not yet support the stronger claim, sometimes implied in vendor and media coverage, that these platforms independently improve hard clinical outcomes - the trial designs needed to demonstrate that are largely still ahead of us. The alarm-fatigue literature offers a useful warning about where this technology could go wrong: a room-level monitoring system that floods staff with low-specificity alerts will be ignored exactly as bedside monitors have been

when poorly tuned. The more promising path, evident in the platforms with the most credible multicenter data so far, treats autonomous monitoring primarily as a measurement and feedback instrument for teams and institutions rather than as a real-time alarm generator, reserving synchronous alerting for a narrow, high-specificity set of events. Anesthesiologists have a direct stake in how this technology matures, both because the anesthesia workstation is one of the richest data sources in the room and because the profession has already lived through one full cycle of automated-monitoring adoption, alarm proliferation, and hard-won alarm-management reform. The lessons from that earlier cycle are directly applicable here, and the field would do well to apply them proactively rather than relearning them room by room.

DECLARATIONS

Acknowledgements

The authors thank the perioperative quality and informatics staff at their respective institutions for discussions that informed this review.

Funding

This review received no external funding.

Conflict of Interest

The authors declare no conflicts of interest relevant to this work.

Author Contributions

AN conceived the review and drafted the initial manuscript. KB performed the literature search and synthesized the sections on team behavior and alarm management. HC critically revised the manuscript for intellectual content and prepared the final version. All authors approved the final manuscript.

Ethics Statement

This is a literature review; no new human or animal data were generated, and no ethics committee approval was required.

Data Availability

No new data were generated for this review. All sources discussed are publicly available in the cited literature.

REFERENCES

1. The Black Box Platform (2023) Surgical Safety Technologies.
2. Artificial intelligence in the operating room: A systematic review of AI models for surgical phase, instruments and anatomical structure identification (2025) PMC12575173.
3. Eight years' experience with automated anesthesia record keeping: lessons learned - new directions taken (1989) International Journal of Clinical

- Monitoring and Computing, DOI:[10.1007/BF01142189](https://doi.org/10.1007/BF01142189).
4. Black box surgery (2022) Stanford Medicine News.
 5. How OR 'Black Boxes' Are Helping to Improve Safety (2023) American Hospital Association Center for Health Innovation Market Scan.
 6. Remote assessment of real-world surgical safety checklist performance using an OR black box: a multicenter study (2023) AHRQ PSNet.
 7. Black Box Technology Shines Light on Improving OR Safety, Efficiency (2023) American College of Surgeons Bulletin, 108(7).
 8. Computer vision in surgery: from potential to clinical value (2022) npj Digital Medicine, DOI:[10.1038/s41746-022-00707-5](https://doi.org/10.1038/s41746-022-00707-5).
 9. From Data to Decision: A Comprehensive Review of Real-Time Analytics and Smart Technologies in the Surgical Suite (2024) Methodist DeBakey Cardiovascular Journal, DOI:[10.14797/mdcvj.1658](https://doi.org/10.14797/mdcvj.1658).
 10. Automated surgical workflow recognition in privacy-preserving depth videos of the operating room (2025) Surgical Endoscopy, DOI:[10.1007/s00464-025-12031-6](https://doi.org/10.1007/s00464-025-12031-6).
 11. Assessing Team Situational Awareness in the Operating Room via Computer Vision (2024) arXiv preprint, arXiv:2406.09185.
 12. Statement on Distractions (2023) American Society of Anesthesiologists Standards and Practice Parameters.
 13. Monitor Alarm Fatigue: An Integrative Review (2012) Biomedical Instrumentation and Technology, 46(4):268-277.
 14. Alarm Fatigue and Patient Safety (2023) Anesthesia Patient Safety Foundation.
 15. Statement on Principles for Alarm Management for Anesthesia Professionals (2023) American Society of Anesthesiologists Standards and Practice Parameters.
 16. Sanchez A, et al. (2024) Artificial intelligence for improving intraoperative surgical care. Global Surgical Education - Journal of the Association for Surgical Education, DOI:[10.1007/s44186-024-00268-z](https://doi.org/10.1007/s44186-024-00268-z).
 17. Artificial intelligence and machine learning approaches for patient safety in complex surgery: a review (2025) PMC12648791.
 18. Integrating Image-Based Artificial Intelligence in the Operating Room: Enhancing Safety and Efficiency While Navigating Legal and Ethical Considerations (2024) Telehealth and Medicine Today.
 19. This AI-powered "black-box" could make surgery safer (2024) MIT Technology Review.
 20. Artificial Intelligence and Robotics in General Surgery: Opportunities and Challenges (2025) PMC13283024.
 21. AI for Quality Assurance in the Operating Room (2026) arXiv preprint, arXiv:2606.30657.