Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2154–2161
https://doi.org/10.54941/ahfe1008441

Artificial Intelligence-Assisted Rapid On-Site Evaluation: A Human Factors-Informed Patient Safety Framework for Cytopathology

Bryan Johnston[1], Jay Kalra[1,2], Heather Ruff[1,3], and Janine Benoit[1,3]

[1]Department of Pathology and Laboratory Medicine, College of Medicine, University of Saskatchewan, Saskatoon, Canada
[2]Royal University Hospital, Saskatchewan Health Authority, 103 Hospital Drive, Saskatoon, Saskatchewan, S7N 0W8, Canada
[3]Saskatoon City Hospital, Saskatchewan Health Authority, 701 Queen St, Saskatoon, Saskatchewan, S7K 0M7, Canada

ABSTRACT

Artificial intelligence (AI) in cytopathology has largely focused on image classification rather than integration into real-time clinical workflows. Rapid on-site evaluation (ROSE), including endobronchial ultrasound-guided transbronchial needle aspiration and endoscopic ultrasound-guided sampling of pancreatic and other lesions, presents a more complex implementation environment. These procedures involve sequential sampling, specimen preparation, adequacy assessment, multidisciplinary communication, and immediate decisions regarding next steps in procedural management. Errors or delays may contribute to unnecessary needle passes, inadequate material, repeat procedures, and potentially delayed diagnoses. Although emerging AI studies demonstrate potential applications in adequacy assessment and malignant cell detection, current ROSE and cytopathology guidance provides limited direction regarding validation, usability, human oversight, or integration into procedural work. To address this gap, we conducted a structured, literature-informed conceptual synthesis of AI-assisted ROSE, procedural cytopathology, digital cytology, and human-factors research. We used this literature to illustrate a purpose-driven approach: define the clinical purpose, determine what would need to be established to judge whether it was achieved, identify the evidence required, and determine what must therefore remain observable before deployment. Three clinical purposes were examined: confirming representative lesional material, preserving and triaging material for molecular or ancillary testing, and determining whether additional sampling is needed. Drawing on these findings, we suggest that the same broad workflow requires different evidence depending on the clinical purpose and that model accuracy alone does not establish safe performance. We therefore recommend treating AI as a component of the procedural cytopathology work system rather than as a stand-alone image classifier. This approach may help move AI-assisted ROSE from isolated image analysis toward safer and more operationally integrated clinical systems.

Keywords: Artificial intelligence, Cytopathology, Workflow integration, Regulation, Prospective evaluation, Human factors

Received September 3, 2026; Revised October 4, 2026; Accepted October 15, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

INTRODUCTION

As artificial intelligence (AI) is integrated into established clinical workflows, the starting point should be the purpose of the workflow rather than the capabilities of a new tool. Integration can change how information is generated, interpreted, communicated, and acted upon, so organizations need to define what the modified workflow is intended to accomplish and what role AI will play in achieving that purpose. Cytopathology provides a useful example. During rapid on-site evaluation (ROSE), cytologic material is assessed while the procedure is underway so the result can inform further sampling and specimen triage (Roy-Chowdhuri et al., 2020). ROSE therefore highlights a broader implementation challenge: after AI changes the workflow, the organization must still be able to determine whether the clinical purpose was achieved.

ROSE also shows why success must be defined in advance. The criterion for sufficient material depends on what the procedure is intended to accomplish. Material adequate for diagnosis may still be insufficient for a complete biomarker workup, and the presence of lesional cells does not by itself establish that further sampling is unnecessary. In one precision-oncology workflow, downstream molecular requirements influenced requests for additional passes and the direction of subsequently acquired material toward cell-block preparation (Al Zoughbi et al., 2021; Hendry et al., 2023). Adequacy is therefore purpose-specific: a specimen may need to be sufficient not only for diagnosis, but also for the ancillary testing required to answer the clinical question. What counts as sufficient may also depend on the local testing workflow and the specimen preparations accepted for those tests. AI is now entering this purpose-defined work. Published AI-ROSE systems can classify cytologic images, detect or segment suspicious cellular material, highlight regions of interest, and produce outputs within time frames potentially relevant to intra-procedural use (Ai et al., 2022; Ashida et al., 2026; Chen et al., 2026). These are important advances, but a timely and accurate AI result does not establish that the modified workflow accomplished its clinical purpose. The material may have been poorly prepared or unrepresentative, the result may be misunderstood or not communicated effectively, or it may arrive too late to inform further sampling or specimen triage.

Human-factors research already establishes that AI should be evaluated as part of the work system in which it operates and that model performance cannot reliably predict how people and AI will perform together (Morey et al., 2025; Salwei and Carayon, 2022). We extend this reasoning by asking what evidence an organization would later need to determine whether an AI-modified workflow achieved a defined clinical purpose. This paper uses AI-assisted ROSE to illustrate a purpose-driven approach: define the clinical purpose, identify what must be established to judge whether it was achieved, and determine what evidence must remain observable before deployment.

METHODS

We conducted a structured, literature-informed conceptual synthesis rather than a systematic review. The source corpus included direct AI-assisted ROSE

studies; professional guidance and empirical studies on ROSE, telecytology, and digital cytology; digital frozen section and intraoperative pathology; and human-factors research on sociotechnical and joint human–AI performance. Telecytology was treated as the closest workflow precedent, while digital frozen section and intraoperative pathology were used as complementary models of time-critical digital interpretation. Direct AI-ROSE studies were compared using a common specimen-to-action pathway: acquisition, preparation, image or field selection, AI output, human interpretation, communication, procedural action, and downstream outcome. We then reasoned backward from three ROSE purposes, confirming representative lesional material, supporting continuation or stopping of sampling, and preserving and triaging material for molecular or ancillary testing, to identify what would need to be established to judge whether each purpose had been achieved and what evidence must therefore remain observable before deployment. The synthesis was intended to illustrate a purpose-driven approach, not to produce a product-specific technical standard or formal evidence-based guideline.

A PURPOSE-DRIVEN APPROACH TO AI-ASSISTED ROSE

This approach starts with the clinical purpose rather than with the available AI capability. Rather than asking where an available AI output might fit, it asks what the workflow is intended to accomplish and what account of performance an organization may later need.
For AI-assisted ROSE, we use four questions:

- What clinical purpose is the AI-modified workflow intended to serve?
- What would need to be established later to judge whether that purpose was achieved?
- What evidence would be needed to make that judgment?
- What must therefore remain observable, linked, and accessible before deployment?

These four questions are not a recommendation to collect every possible data element. The evidence required will depend on the purpose, workflow, and context.

APPLYING THE APPROACH TO AI-ASSISTED ROSE

ROSE provides a useful clinical demonstration of the approach because the same broad workflow can serve different purposes. A specimen may be acquired, prepared, assessed, communicated, and acted upon in a similar sequence, yet the evidence needed to judge performance changes with the clinical question (Al Zoughbi et al., 2021; Hendry et al., 2023; Roy-Chowdhuri et al., 2020). The starting point should therefore not be, What can the AI detect? It should be, What would we need to establish to know that the AI-modified workflow accomplished its intended purpose? Three examples show how the intended purpose selects the relevant evidence

and performance measures. Figure 1 depicts AI-assisted ROSE as a feedback loop linking specimen acquisition, preparation, AI assessment, human interpretation, communication, and procedural action. The same broad workflow can support different clinical purposes, which determine which aspects of performance need to remain observable.

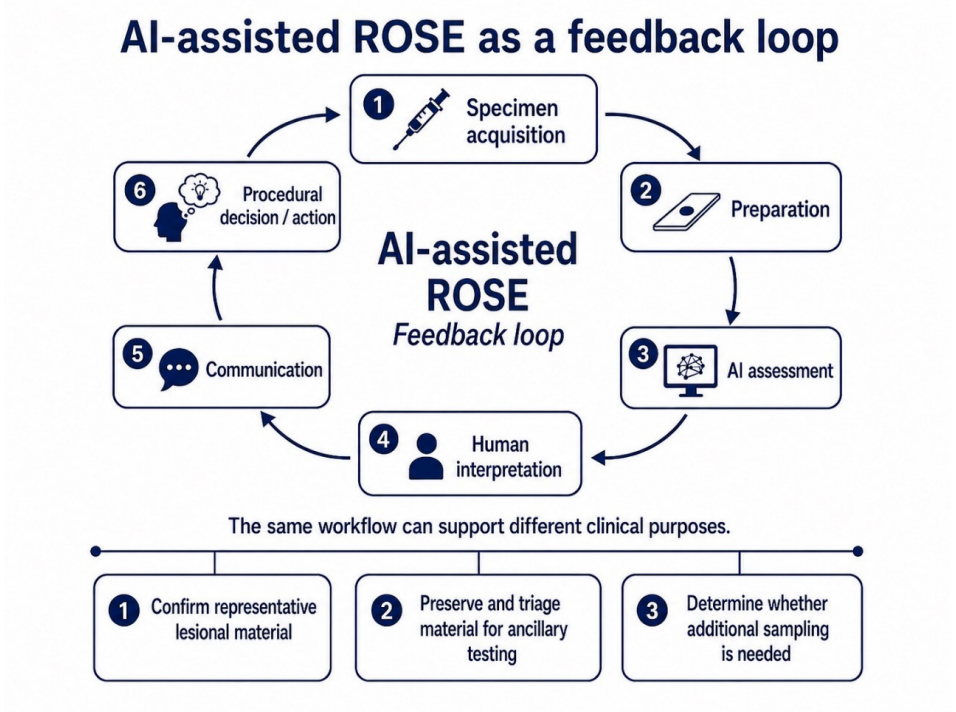


Figure 1: AI-assisted ROSE as a feedback loop. Different clinical purposes require different evidence from the same workflow.

Confirming That Representative Lesional Material Has Been Obtained

One purpose of ROSE is to determine whether the procedure has sampled the intended lesion. Professional guidance describes ROSE as providing real-time feedback about whether representative material is being obtained and allowing that information to be communicated while further sampling remains possible (Roy-Chowdhuri et al., 2020). An AI system could contribute by detecting cellular material, identifying suspicious cell groups, or highlighting regions for human review. Existing AI-ROSE studies demonstrate high technical performance and, in some systems, rapid output for these bounded tasks (Ai et al., 2022; Ashida et al., 2026; Tian et al., 2025; Zhang et al., 2022).

However, image-level performance does not by itself establish that representative material was obtained. Both human and AI assessment depend on the material prepared and presented for review; neither directly evaluates the entire sampled specimen. Depending on implementation, the

AI may assess a whole-slide image or selected fields, and the preparation may not represent the full pass or intended target. To evaluate representative sampling, an organization may therefore need to link the patient, lesion, site and pass with the material presented to AI, the AI output, the human response, and any subsequent sampling decision. Relevant measures might include false reassurance, unnecessary additional passes, missed lesional material, and eventual diagnostic yield. Evaluating representative sampling therefore requires more than the AI result alone; it requires evidence linking the sampled material to the human interpretation and subsequent procedural response.

Supporting a Decision to Continue or Stop Sampling

A second, downstream purpose is to support decisions about whether sampling should continue. ROSE can inform the number of passes and, once adequate material has been obtained for the relevant clinical goals, may support a decision to stop (Roy-Chowdhuri et al., 2020). When AI contributes to this decision, its role is no longer limited to highlighting suspicious material; its output may influence whether the procedural opportunity for further sampling remains open. Evaluating this use therefore requires more than knowing whether the AI labelled an image "adequate." An organization may need to establish what adequacy meant for that procedure, what material was assessed, who could view the AI result, how human users interpreted the AI result, how uncertainty or disagreement was handled, what was communicated to the proceduralist, and why sampling continued or stopped. Timing is also part of performance: a correct result that arrives after the decision has been made cannot serve its intended clinical purpose.

Human-factors research shows that model accuracy does not reliably predict how people and AI will perform together, and misleading AI can degrade joint performance (Morey et al., 2025). In a stop-or-continue context, false adequacy is therefore a safety-critical failure because it may contribute to premature stopping before sufficient material has been obtained. Relevant measures could include premature stopping, unnecessary additional passes, disagreement or override, time from specimen acquisition to communicated adequacy assessment, downstream specimen sufficiency, and repeat procedures. An AI validated to identify malignant-appearing cells should not automatically be treated as validated to determine that sampling can stop, particularly when stopping depends on diagnostic and downstream testing goals. The task for which the AI was validated and the procedural authority assigned to its output should remain aligned.

Preserving and Triaging Material for Molecular or Ancillary Testing

A third purpose is to support obtaining and appropriately allocating enough material for downstream testing. This makes the limits of a generic adequacy label particularly clear. Material sufficient to establish a diagnosis may still be insufficient for a complete biomarker workup. Hendry and colleagues found high diagnostic adequacy among thoracic specimens obtained with ROSE, while some specimens submitted for

predictive biomarker testing could not support the complete workup, often because of the greater cellular requirements of PD-L1 testing (Hendry et al., 2023). The downstream purpose can also influence what happens during the procedure. In the precision-oncology workflow described by Al Zoughbi and colleagues, inadequate fine-needle aspiration material prompted additional passes, while further material could be directed toward cell-block preparation after tumour was identified. The real-time adequacy assessment was strongly associated with subsequent molecular-profiling success (Al Zoughbi et al., 2021). For this purpose, adequacy may depend not only on the amount of material but also on tumour cellularity, the tests required, and the specimen preparations or aliquots accepted in the local workflow. An AI system used for this purpose might help identify tumour-rich material or contribute to an adequacy assessment, but its performance should not be judged only by agreement with a cytologic diagnosis. To determine whether the workflow accomplished its downstream purpose, an organization may need to link the intended test and adequacy criterion with the material assessed, the AI and human assessment, subsequent sampling or specimen allocation, and the eventual testing outcome. Relevant measures could include downstream testing adequacy, failure because of insufficient material, avoidable tissue depletion, and repeat biopsy.

The same broad workflow therefore needs to be judged differently depending on the clinical purpose. Confirming lesional material, deciding whether further sampling is needed, and preserving and triaging material for ancillary testing are not interchangeable claims of "AI-ROSE performance." By defining the purpose before deployment, an organization can identify which aspects of the ROSE workflow need to remain observable for evaluation, discrepancy review, and quality improvement. Table 1 summarizes the three purpose-specific performance accounts and the evidence that may need to remain observable.

Table 1: Illustrative purpose-specific accounts for AI-assisted ROSE.

Clinical Purpose	What Must Be Established?	What Evidence may Need to Remain Available?
Confirm representative lesional material	Was the intended lesion represented?	Lesion, site, and pass; material assessed; AI and human assessment
Determine whether additional sampling is needed	Were the relevant sampling and testing goals sufficiently met to justify stopping?	Adequacy criterion; material assessed; communication and timing; final specimen sufficiency
Preserve and triage material for molecular or ancillary testing	Did sampling and specimen allocation provide enough material for the intended downstream test?	Intended test; adequacy criterion; additional sampling or allocation; final testing outcome

DISCUSSION

AI-assisted ROSE makes the distinction between model performance and clinical workflow performance particularly clear. Intended-use and workflow approaches define what AI is meant to do and how it will participate in clinical work, while human-factors research emphasizes evaluating people and AI together within the wider work system (Morey et al., 2025; Salwei and Carayon, 2022). In AI-assisted ROSE, we therefore ask a prospective question: what evidence would an organization later need to determine whether the AI-modified workflow accomplished its clinical purpose? These examples show that AI-assisted ROSE should be evaluated against the clinical purpose being supported. Representative sampling, decisions about further sampling, and ancillary-testing sufficiency require different evidence from the same broad workflow. The purpose therefore determines which aspects of the workflow need to remain observable for local evaluation and quality improvement.

Telecytology provides the closest workflow precedent because it retains the key procedural, communicative, and sampling relationships of ROSE (Lin, 2018; Lin et al., 2024). Digital frozen section and intraoperative pathology provide complementary examples of time-critical digital interpretation (Zheng et al., 2025). AI adds another layer because its output must be understood in relation to human interpretation, communication, and subsequent procedural action. Whether this purpose-driven approach is useful beyond ROSE should be tested in other time-critical clinical workflows.

CONCLUSION

In AI-assisted ROSE, the clinical purpose should guide both the role assigned to AI and the evidence needed to evaluate the modified workflow. Model accuracy alone does not establish safe performance. Evaluation must connect the specimen, AI assessment, human interpretation, communication, and procedural action to determine whether the intended purpose was achieved.

ACKNOWLEDGMENT

The authors would like to acknowledge that the material in this paper forms part of a thesis to be submitted by Bryan Johnston to the University of Saskatchewan in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

REFERENCES

Ai, D., Hu, Q., Chao, Y.-C., Fu, C.-C., Yuan, W., Lv, L., Ye, D., Li, C., Ye, M., Zhang, Y., Hong, Q., Hu, J., Xu, X., Zhang, L., Jiang, Q., Wang, X., Fang, Q., Wang, B., Hou, Y., Zhang, X., 2022. Artificial intelligence-based rapid on-site cytopathological evaluation for bronchoscopy examinations. Intell.-Based Med. 6, 100069. https://doi.org/10.1016/j.ibmed.2022.100069

Al Zoughbi, W., Kim, D., Alperstein, S.A., Ohara, K., Manohar, J., Greco, N., Khani, F., Robinson, B.D., Rao, R.A., Elemento, O., Mosquera, J.M., Siddiqui, M.T., 2021. Incorporating cytologic adequacy assessment into precision oncology workflow using telepathology: An institutional experience. Cancer Cytopathol. 129, 874–883. https://doi.org/10.1002/cncy.22441

Ashida, R., Kuwahara, T., Koshikawa, T., Hashimoto, K., Okuno, N., Haba, S., Kawaji, Y., Tamura, T., Yamashita, Y., Itonaga, M., Kiriyama, Y., Yamao, K., Hara, K., Kitano, M., 2026. AI -Assisted Real-Time Cytologic Diagnosis During EUS - FNA of Pancreatic Masses (With Video). Dig. Endosc. 38, e70081. https://doi.org/10.1111/den.70081

Chen, Y., Du, C., Gu, P., Li, C., Meng, H., Kong, F., 2026. Feasibility of artificial intelligence-based rapid on-site evaluation for the diagnosis of pulmonary disease. Front. Oncol. 16, 1836366. https://doi.org/10.3389/fonc.2026.1836366

Hendry, S., Mamotte, L., Mesbah Ardakani, N., Leslie, C., Tesfai, Y., Grieu-Iacopetta, F., Izaac, K., Singh, S., Ardakani, R., Thomas, M., Giardina, T., Robinson, C., Frost, F., Amanuel, B., 2023. Adequacy of cytology and small biopsy samples obtained with rapid onsite evaluation (ROSE) for predictive biomarker testing in non-small cell lung cancer. Pathology (Phila.) 55, 917–921. https://doi.org/10.1016/j.pathol.2023.08.002

Lin, O., 2018. Telecytology for rapid on-site evaluation: current status. J. Am. Soc. Cytopathol. 7, 1–6. https://doi.org/10.1016/j.jasc.2017.10.002

Lin, O., Alperstein, S., Barkan, G.A., Cuda, J.M., Kezlarian, B., Jhala, D., Jin, X., Mehrotra, S., Monaco, S.E., Rao, J., Saieg, M., Thrall, M., Pantanowitz, L., 2024. American Society of Cytopathology Telecytology validation recommendations for rapid on-site evaluation (ROSE). J. Am. Soc. Cytopathol. 13, 111–121. https://doi.org/10.1016/j.jasc.2023.12.001

Morey, D.A., Rayo, M.F., Woods, D.D., 2025. Empirically derived evaluation requirements for responsible deployments of AI in safety-critical settings. Npj Digit. Med. 8, 374. https://doi.org/10.1038/s41746-025-01784-y

Roy-Chowdhuri, S., Dacic, S., Ghofrani, M., Illei, P.B., Layfield, L.J., Lee, C., Michael, C.W., Miller, R.A., Mitchell, J.W., Nikolic, B., Nowak, J.A., Pastis, N.J., Rauch, C.A., Sharma, A., Souter, L., Billman, B.L., Thomas, N.E., VanderLaan, P.A., Voss, J.S., Wahidi, M.M., Yarmus, L.B., Gilbert, C.R., 2020. Collection and Handling of Thoracic Small Biopsy and Cytology Specimens for Ancillary Studies: Guideline From the College of American Pathologists in Collaboration With the American College of Chest Physicians, Association for Molecular Pathology, American Society of Cytopathology, American Thoracic Society, Pulmonary Pathology Society, Papanicolaou Society of Cytopathology, Society of Interventional Radiology, and Society of Thoracic Radiology. Arch. Pathol. Lab. Med. 144, 933–958. https://doi.org/10.5858/arpa.2020-0119-CP

Salwei, M.E., Carayon, P., 2022. A Sociotechnical Systems Framework for the Application of Artificial Intelligence in Health Care Delivery. J. Cogn. Eng. Decis. Mak. 16, 194–206. https://doi.org/10.1177/15553434221097357

Tian, Y., Zhang, J., Rao, X., Zhou, B., Zhao, X., He, Y., Sheng, H., Nian, W., Cai, Y., Rong, L., 2025. AI-assisted rapid on-site evaluation ROSE of EUS-FNA cytopathology for pancreatic solid lesions: A two-stage deep learning approach. Endosc. Ultrasound 14, 314–327. https://doi.org/10.1097/eus.0000000000000154

Zhang, S., Zhou, Y., Tang, D., Ni, M., Zheng, J., Xu, G., Peng, C., Shen, S., Zhan, Q., Wang, X., Hu, D., Li, W.-J., Wang, L., Lv, Y., Zou, X., 2022. A deep learning-based segmentation system for rapid onsite cytologic pathology evaluation of pancreatic masses: A retrospective, multicenter, diagnostic study. eBioMedicine 80, 104022. https://doi.org/10.1016/j.ebiom.2022.104022

Zheng, X., Zheng, K., Wen, Y., Meng, J., Zhang, X., Wen, X., Zhao, Z., Zheng, C., Cai, X., Lin, J., Chen, J., Duan, J., Jiang, L., Yuan, W., Li, X., Xie, D., Cai, Y., Zhang, J., Cai, M., 2025. An end-to-end multifunctional AI platform for intraoperative diagnosis. Npj Digit. Med. 8, 460. https://doi.org/10.1038/s41746-025-01808-7