AI in Healthcare: The Risk of Insufficient Governance

James Ritter

University of San Diego

HCIN 559

Dr. Brenda Boone

24 June 2024

AI in Healthcare: The Risk of Insufficient Governance

Integrating Artificial Intelligence (AI) into the healthcare industry provides seemingly boundless opportunities to revolutionize patient care. While AI-based technologies promise a paradigm shift in clinical care that enhances the speed of healthcare delivery, optimizes clinical decision support, and facilitates personalized care, the efficacy of these tools is often overestimated due to inherent technological limitations (Olawade et al., 2024). The output of an AI system is only as good as the data it receives—garbage in, garbage out. Furthermore, an AI system depends entirely on complex algorithms often abstracted away from users, providing significant barriers to process modification and user validation (ECRI, 2024). Barring the excitement surrounding this burgeoning technology, the responsible implementation of AI systems necessitates robust governance that validates such systems' efficacy, quality, and reliability in clinical settings.

AI is a comprehensive term encompassing a class of analytic methodologies that implement sophisticated statistical techniques via computational algorithms to provide deeper insights into a dataset. Different frameworks offer various results: while supervised machine learning can help predict outcomes, unsupervised models are more adept at identifying patterns and simplifying datasets. These methodologies can be compounded into more complex neural networks that compute these analyses automatically. When trained on large amounts of high-quality data, these technologies can provide accurate and insightful results (IBM, 2023).

One of the significant challenges in implementing AI in healthcare is the quality and standardization of real-world datasets. Many of these datasets are erroneous, biased, and often incomplete, requiring extensive data cleaning and compressing before they can be used with AI-based technologies. The healthcare industry lacks standardization in data collection, formatting, and storage, with each healthcare organization dictating the structure and management of patient data (Giard & Enlitic, 2023). This lack of standardization can lead to unexpected responses from AI models trained on one organization's data when exposed to a different organization's data structure. Additionally, large-scale, inter-organizational data-sharing agreements are never guaranteed, limiting the availability of high-quality data for model training and testing. AI algorithms built upon inconsistent and low-quality datasets will likely produce inconsistent and low-quality results. (Olawade et al., 2024).

Compounding this issue, AI algorithms are often abstracted behind layers of user interfaces. While this enhances the technology's accessibility to users from diverse academic backgrounds, it creates a transparency issue. As a result, average users have little to no understanding of, or ability to validate, the mechanisms that generate results. Ennab and Mcheick (2022) liken this deficit in interpretability to a "black box," stating that "building trust in the machine-learning models has become a prerequisite for the ultimate adoption of AI systems [in healthcare]" (p. 2). As AI systems become increasingly integrated in clinical settings, ensuring that algorithmic transparency and clinician engagement are at the forefront of future development is imperative.

While revolutionary in its promise to improve the efficiency and quality of patient care, inherent deficiencies in AI algorithm and system development require significant oversight. With the rise of these technologies in clinical settings, comprehensive and robust governance programs that synchronize intra- and inter-organizational data collection, regulate AI algorithm implementation, ensure transparency, and facilitate technical education are crucial. This white paper aims to advocate and provide recommendations for widespread governance organizations that manage the integration of AI-based technologies in the healthcare industry.

Literature Review

While AI-based technologies and their applications in the healthcare industry have garnered significant excitement, their integration into safety- and quality-minded clinical care organizations poses substantial challenges. As these organizations aim to improve clinical efficiency by deploying AI-based technologies, inherent technological deficiencies that threaten patient safety may become apparent. This review aims to provide a comprehensive overview of the barriers to adoption and potential shortcomings of AI-based technologies in the health sector.

A systematic review by Anugerah & Hidayanto (2023) aimed to identify the factors associated with the acceptance of AI-based technologies in patient care. Through a review of 27 papers, the researchers identified thirteen individual factors that affected AI-based tools in diagnostics and clinical decision support (CDS). These factors were specified using the Technology, Organization, People, Environment (TOPE) framework, which provides a peer-reviewed approach to qualify the acceptance of new technologies in organizational settings. The researchers identified usability as the most important technological factor for AI-based tools in both diagnostics and CDS applications. Standardization was designated a crucial organizational factor, while trust and ease of use were identified as valuable people factors. These findings exemplify the technical, social, and organizational underpinnings of this technology that must be considered for acceptance in clinical settings.

Importantly, AI-based tools must provide insight into patient care that improves the quality and reliability of diagnostics. A meta-analysis by McGenity et al. (2023) aimed to quantify the accuracy of AI in digital pathology. This systematic review included 48 studies implementing any AI image-detection algorithm on pathology samples of varying disease types, amounting to 152,000 diagnostic images in total. While the authors noted limitations due to the heterogeneity of the samples and potential publication biases, sensitivity and specificity scores across all studies were approximately 96% and 93%, respectively. Thus, these AI-based tools provided false negative results in approximately 4% of cases and reported false positives in approximately 7%. The authors noted that sensitivity and specificity are drastically improved in studies that include a variety of data sources, underscoring the need for variability and accuracy for any AI training dataset. While these preliminary levels of accuracy are promising, robust AI-based diagnostic tools will require significant development to improve diagnostic specificity.

Unfortunately, the complex algorithms that form the foundation of a successful AI system are often abstracted away through layers of user interfaces. When these algorithms inevitably produce erroneous or otherwise questionable results, this significant abstraction might pose a barrier to developing or maintaining a clinician's trust in the AI. Olawade et Al. (2022) explored this and other considerations, denoting the significant training and workflow integrations clinicians will need to incorporate AI-based technologies in practice effectively. The researchers also distinguished between AI-based technologies that assist in clinical care versus administrative applications: while an ophthalmologist may benefit from an AI-enhanced image-detection algorithm, primary care providers may enjoy automating specific tasks through AI-improved clinical decision support. These specialties were identified as most likely to immediately benefit from incorporating AI-based tools in clinical settings. Diagnostic tools would likely require the most scrutiny, with patient safety standards demanding significant regulatory approval to qualify the algorithm's accuracy, validity, reliability, and performance. These considerations should be at the forefront of any plan to implement AI-based technologies at the technical and organizational levels.

In planning for the inevitable integration of powerful AI-based tools in clinical settings, healthcare organizations must balance the perceived benefits these tools may provide with the challenges associated with adopting disruptive technologies. Jongsma, Sand, and Milota challenge the notion that AI-based tools will improve efficiency in healthcare,

> "...the confusion of accuracy for efficiency in terms of workload reduction is flawed because it rests on a too narrow assessment of what constitutes workload in the first place...for example AI systems for radiology...will require a constant stream of expert-annotated images to maintain system accuracy. If these annotations must be completed separately...from standard annotation processes...this additional labor will have to be factored into clinicians' already heavy workload..." (2024, p. 1).

Alongside the additional workload prescribed to clinicians providing diagnostic training data for the AI, providers must also contend with the reality of imperfect technology. The authors noted that each clinician must find a delicate balance between trusting and scrutinizing the results of an AI-powered diagnostic tool since these technologies are prone to making mistakes, especially when fed erroneous training data. These mistakes are accompanied by significant patient safety concerns and the additional burden placed on clinicians to validate an AI's response on the fly.

Ultimately, the literature indicates that while promising, AI-based technologies in the healthcare industry are still early in their development and integration. Although these algorithms may provide high diagnostic accuracy, significant technical development is required to ensure these tools have consistent access to variable and accurate data. The development cycle must be accompanied by clinician involvement aimed at training and building trust with the technology to achieve efficiency promises and guarantee patient safety.

Identified Solution

As healthcare organizations begin implementing AI-based technologies in clinical practice, expansive governance programs must be in place to maintain quality and reliability in patient care. This white paper will suggest a comprehensive governance program that coordinates data management and AI technology integration at the organizational and governmental levels.

This proposed governance program must set forth a framework by which healthcare organizations can facilitate the development of AI-based technologies in clinical settings that guarantee reliability and safety in patient care. Ovregaard et al. (2023) propose a methodology that incorporates the principles of a Quality Management System (QMS) that utilizes extensive technical and procedural documentation, continual monitoring, and effective risk management to ensure healthcare organizations (and their AI partners) promote a culture of quality, reliability, transparency, progress, and compliance. The authors champion a system of risk identification, enumeration, mitigation, and monitoring to ensure that healthcare organizations and their partners have a reliable mechanism for developing and implementing future iterations of the technology. A robust QMS-based platform like this would allow healthcare organizations and their partners to implement AI-based technologies systematically and in a quality-focused manner that promotes reliable healthcare delivery and patient safety.

In addition to the regulatory environment facilitated by the QMS framework defined above, the governance program will also encourage healthcare organizations to implement technological infrastructure upgrades to support AI-enabled Electronic Health Records (EHRs). Schuessler et al. (2022) note the importance of high-quality data to reduce bias in AI training models and improve predictive accuracy. Additionally, the authors recommend implementing interoperability resources via the Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) protocol to facilitate the transfer of standardized medical data to assist in algorithm refinement and further AI model development. The proposed governance program must incentivize healthcare organizations to implement data quality interventions, staff training, and EHR upgrades to ensure that providers gather complete and accurate clinical data for every patient. Furthermore, the governance program must facilitate and regulate HL7 FHIR data transfer between organizations. This may be through the introduction of a secure, accessible, and efficient platform by which healthcare organizations and providers can transfer depersonalized medical data in a standard format.

The proposed governance program would facilitate the adoption of highly effective and efficient AI systems through quality-minded interventions that affect the technical, personal, and organizational culture around data collection, technical infrastructure, interoperability, and, ultimately, reliable and safe patient care.

Failure Mode Effect Analysis

To mitigate potential reliability and safety concerns, this white paper will provide a Failure Mode Effect Analysis (FMEA) to study the development and integration of AI-based healthcare delivery tools across five main steps:

1. Health IT Organizations (AI development partners) produce innovations in Health AI and submit preliminary findings to the proposed governance program's Institutional Review Board (IRB). See Figure A1(1) for a detailed overview of the process and Table A1 for further analysis.
2. Once approved, healthcare organizations can update software and hardware according to the developers' specifications. See Figure A1(2) for a detailed overview of the process and Table A2 for further analysis.
3. Once software and hardware functionalities have been validated, healthcare organizations can implement these new tools in patient care settings. See Figure A2(3) for a detailed overview of the process and Table A3 for further analysis.
4. Healthcare organizations and the governance program must begin monitoring the tool for quality and reliability. See Figure A2(4) for a detailed overview of the process and Table A4 for further analysis.
5. Healthcare organizations and providers must report errors and other patient safety concerns when they occur. The governance program's IRB must review these error reports and determine a course of action that upholds patient safety standards and reliable healthcare delivery. See Figure A3(5) for a detailed overview of the process and Table A5 for further analysis.

Quality Measurement Plan

The governance program introduced in this white paper aims to harness the potential of AI-based technologies to enhance patient care while ensuring safety and reliability. The program, underpinned by a comprehensive quality measurement plan (QMP), is designed to ensure and promote highly reliable clinical care. With patient safety and comfort as the top priority, this QMP will establish a robust framework for reporting and responding to any shortcomings of AI-based technologies. The success of any AI-based technology in clinical practice will be measured by its algorithmic accuracy and clinician buy-in.

At the heart of the governance program, the QMS system is pivotal in maintaining and monitoring an organization's AI infrastructure. It facilitates accurate and timely data collection procedures, ensuring a robust foundation for quality reporting. Staff will be surveyed on the ease of use, effectiveness, and accuracy of each tool used in patient care, with their responses quantified to reflect the quality of care and account for any potential biases. Additionally, inline quality control tools will allow clinicians to report accuracy concerns, offering valuable insights into the robustness of any AI model developed for clinical care.

The governance program will maintain consistent monitoring practices to aggregate and review quality measures for tools actively used across participating organizations. This institutional review board will consist of representatives from diverse professional and academic backgrounds, providing insight into patient care, the health sciences, and computer technology. Significant errors and patient safety concerns will require immediate attention from this board and regular monitoring. The diversity of the monitoring board and its ability to respond to quality and safety concerns will provide valuable metrics regarding the effectiveness of the governance program.

Data will be collected and aggregated in near-real time, barring any technological limitations on the central QMS platform. The platform will require robust data infrastructure technologies to ensure accurate, secure, and timely data transfer in a standardized and accessible format. Appropriate data management technologies, alongside consistent and standardized data analytics tools, should provide accurate insight into this large dataset for quality monitoring. Data transfer latency and platform up-time will provide valuable metrics regarding the effectiveness of the QMS system itself.

This QMP provides a framework that gauges the quality of various AI-based tools through a QMS system that facilitates clinician reporting and inline accuracy evaluation. The governance program aims to quantify and maintain high reliability and safety while incorporating these powerful and promising technologies through robust data infrastructure, analysis, and regular monitoring.

Conclusion

As healthcare organizations begin to incorporate various AI-based technologies into clinical practice, it is imperative to introduce a comprehensive oversight program that upholds the quality and reliability of patient care while monitoring the incorporation of these revolutionary technologies. Issues with the accessibility of complete, accurate, and structured data, alongside concerns over algorithmic complexity and abstraction, may prevent clinician buy-in and potentially increase medical errors. While developments in this industry are promising in their potential to overhaul the diagnostic and clinical decision-making process, various technical and organizational limitations must be accounted for to promote safe and reliable patient care.

This white paper proposes the formation of a central governance program that promotes patient safety through consistent monitoring, reporting, and reviewing. We suggest the incorporation of a quality management system tasked with managing the incorporation and performance of these technologies, managed closely by a diverse institutional review board of established professionals with a vested interest in promoting patient safety alongside the advancement of Health AI.

References

Anugerah, G. B., & Hidayanto, A. N. (2023). What factors contribute to the acceptance of artificial intelligence in the healthcare sector: A systematic review. In 2023 3rd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA) (pp. 477-482). Denpasar, Bali, Indonesia. https://doi.org/10.1109/ICICyTA60173.2023.10428744

Emergency Care Research Institute (ECRI). (2024). ECRI 2024 top 10 health technology hazards executive brief [White paper].

Ennab, M., & Mcheick, H. (2022). Designing an interpretability-based model to explain the artificial intelligence algorithms in healthcare. Diagnostics, 12, 1557. https://doi.org/10.3390/diagnostics12071557

Giard, C., & Enlitic. (2023, December 15). Revolutionizing healthcare through data standardization | HIMSS. Www.himss.org. https://www.himss.org/resources/revolutionizing-healthcare-through-data-standardization

International Business Machines Corporation (IBM). (2023, June 3). AI vs. machine learning vs. deep learning vs. neural networks. https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

Jongsma, K. R., Sand, M., & Milota, M. (2024). Why we should not mistake accuracy of medical AI for efficiency. NPJ Digital Medicine, 7(1), 1–3. https://doi.org/10.1038/s41746-024-01047-2

Mcgenity, C., Clarke, E., Jennings, C., Matthews, G., Cartlidge, C., Freduah-Agyemang, H., Stocken, D., & Treanor, D. (2023). Artificial intelligence in digital pathology: a diagnostic test accuracy systematic review and meta-analysis. https://arxiv.org/pdf/2306.07999

Olawade, D. B., David-Olawade, A. C., Wada, O. Z., Asaolu, A. J., Adereni, T., & Ling, J. (2024). Artificial intelligence in healthcare delivery: Prospects and pitfalls. Journal of Medicine, Surgery, and Public Health, 3(100108). https://doi.org/10.1016/j.glmedi.2024.100108.

Overgaard, S. M., Graham, M. G., Brereton, T., Pencina, M. J., Halamka, J. D., Vidal, D. E., & Economou-Zavlanos, N. J. (2023). Implementing quality management systems to close the AI translation gap and facilitate safe, ethical, and effective health AI solutions. Npj Digital Medicine, 6(1), 1–5. https://doi.org/10.1038/s41746-023-00968-8

Schuessler, M., Bärnighausen, T., & Jani, A. (2022). Organisational Readiness for the Adoption of Artificial Intelligence in Hospitals. In M. Corrales Compagnucci, M. L. Wilson, M. Fenwick, N. Forgó, & T. Bärnighausen (Eds.), AI in eHealth: Human Autonomy, Data Governance and Privacy in Healthcare (pp. 334–377). chapter, Cambridge: Cambridge University Press.

Appendix A

Figure A1

FMEA Process Flow (Steps 1 & 2)

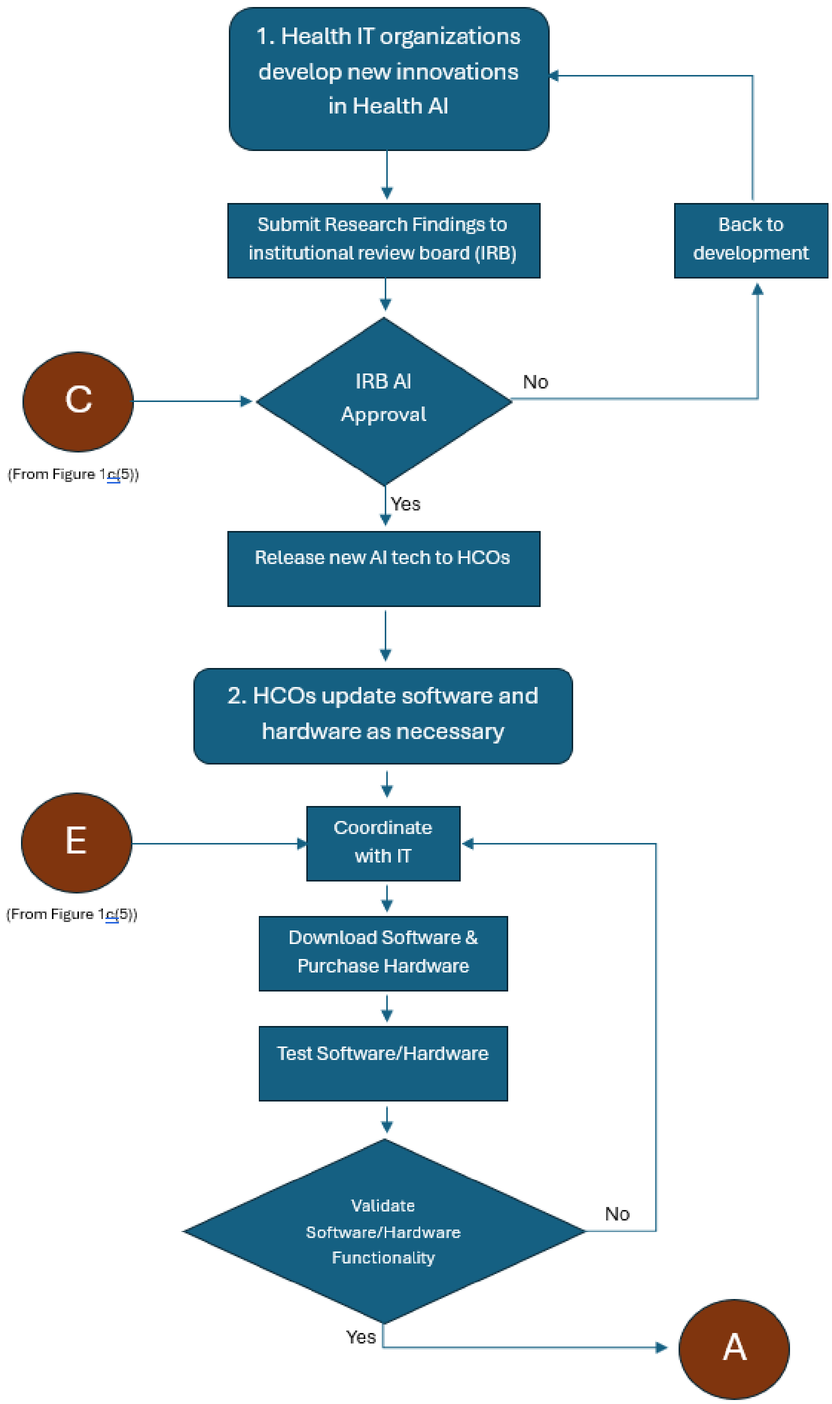

Figure A2

FMEA Process Flow (Steps 3 & 4)

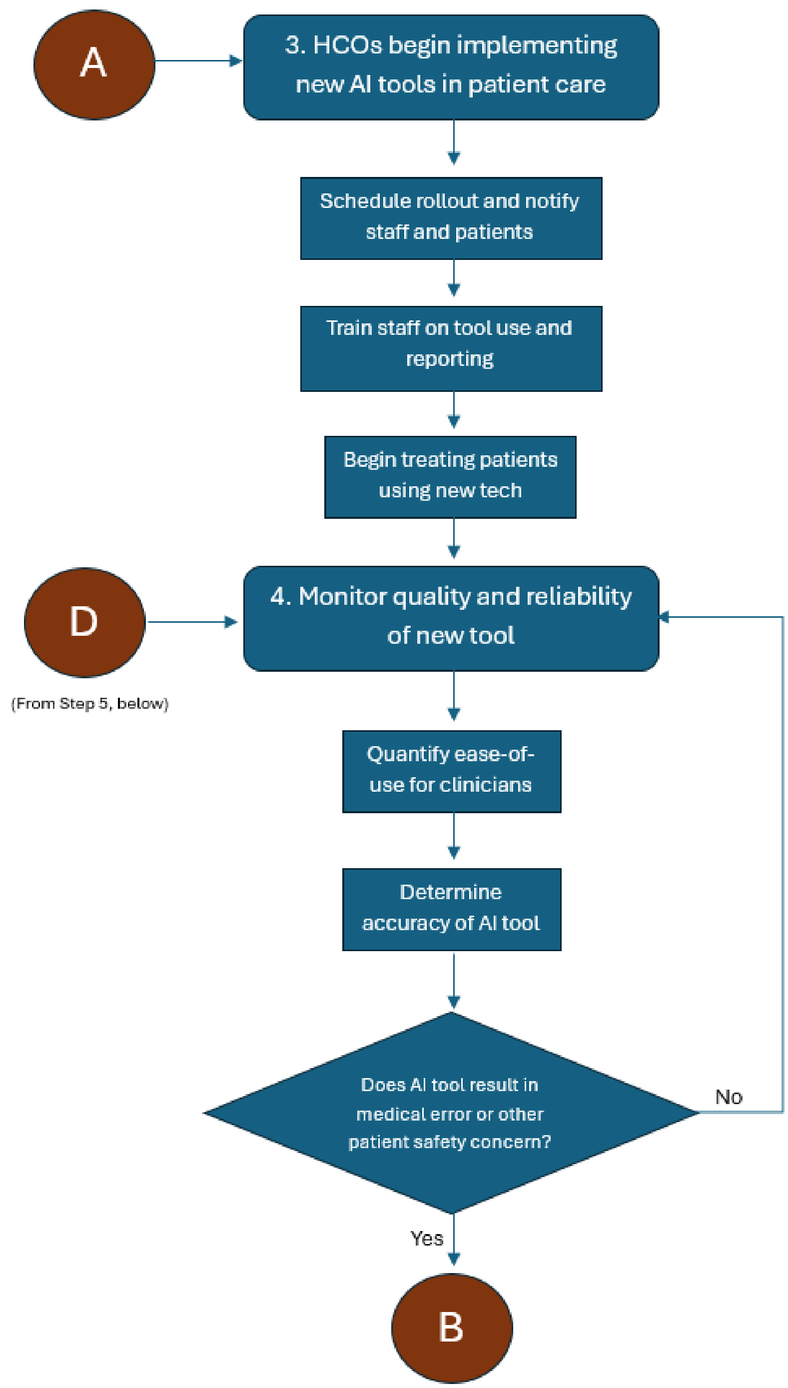

Figure A3

FMEA Process Flow (Step 5)

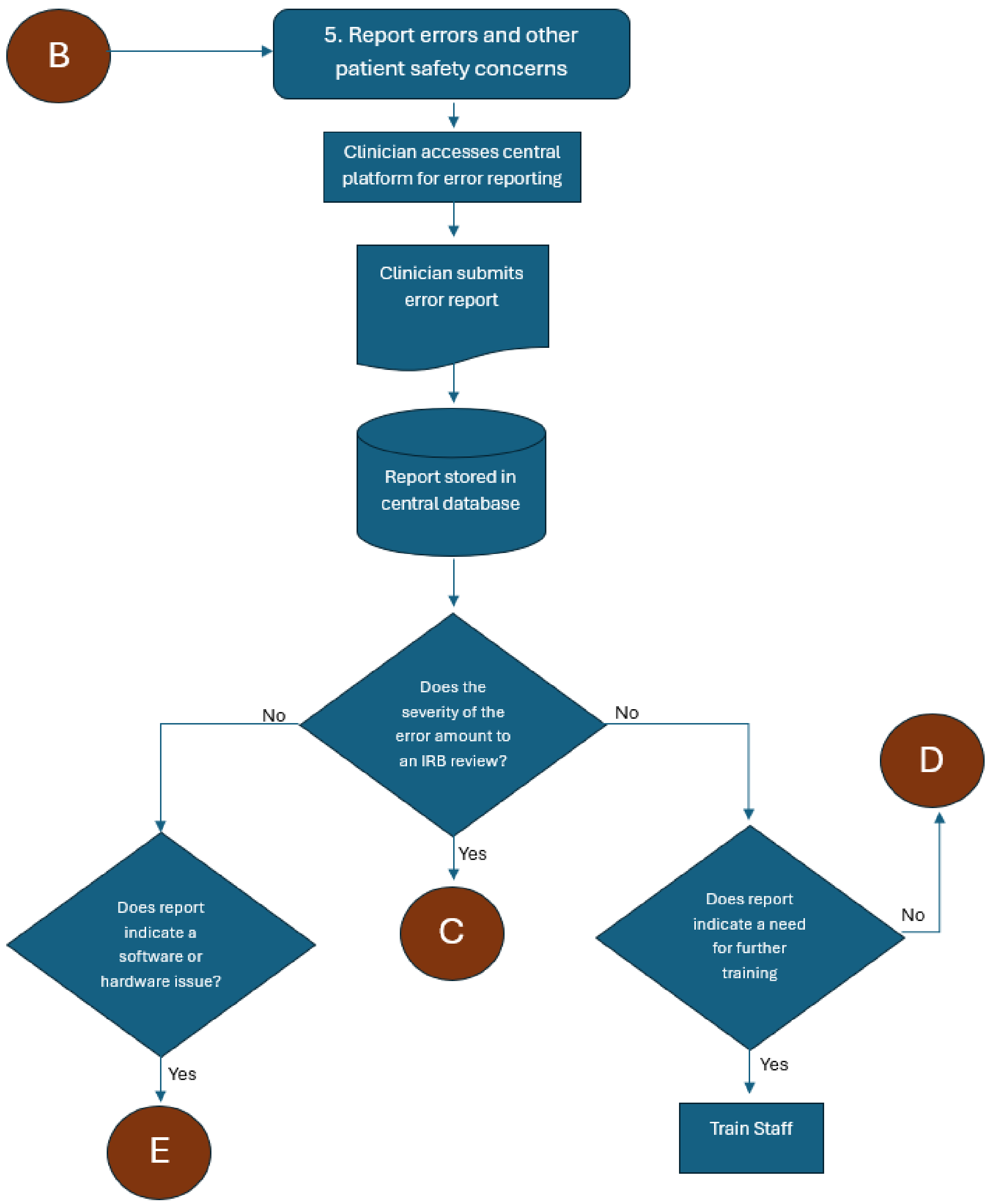

Table A1

FMEA Process Step 1 – Health IT Develops New AI Tool

Process Step #1	1	Process Step	Health IT Develops new AI tool		
	2	Potential Failure Mode	The AI model is overtrained/inaccurate	AI model uses the wrong statistical model	AI model integrated into faulty software
	3	Potential Cause	A small amount and low variety of training data	Wrong statistical assumptions made during development	Improper programming techniques
	4	Severity	Major	Major	Moderate
	5	Probability	Frequent	Uncommon	Uncommon
	6	Hazard Score	12	6	4
	7	Action	Eliminate	Eliminate	Control
	8	Description of Action	1. Ensure the model has access to variable data 2. Use actual data to develop a model 3. Check for overfitting using standard ML/AI methods	1. Validate the statistical method using test data 2. Remove the model from patient care to prevent severe error 3. Report findings to IRB and developer	1. Report software bugs as soon as they are found 2. Enroll the IT department to communicate faults with the developer 3. Coordinate with developers to maintain software patches

Table A2

FMEA Process Step 2 – HCO Software and Hardware Updates

Process Step #2	1	Process Step	HCO Software and Hardware Updates		
	2	Potential Failure Mode	Software update does not complete properly	Technical Infrastructure cannot support software	Updates cause unforeseen security issues
	3	Potential Cause	Insufficient technical infrastructure	Insufficient computing power	Faulty software design
	4	Severity	Minor	Moderate	Moderate
	5	Probability	Occasional	Occasional	Occasional
	6	Hazard Score	3	6	6
	7	Action	Eliminate	Eliminate	Eliminate
	8	Description of Action	1. Plan software updates during downtime 2. Dedicate IT resources to ensure software updates properly 3. Enlist the help of developers to ensure the software update completes	1. Check hardware requirements before initiating the upgrade 2. Upgrade infrastructure to ensure minimum hardware requirements 3. Ensure IT staff are available during hardware upgrades and software updates	1. Ensure cybersecurity systems are in place to monitor vulnerabilities 2. Deploy software updates in a siloed testing environment before deploying to production 3. Employ penetration testing to ensure that software updates do not introduce unforeseen vulnerabilities

Table A3

FMEA Process Step 3 – HCO Implements New AI Tool

Process Step #3	1	Process Step	HCO Implements New AI Tool		
	2	Potential Failure Mode	Staff reject AI tool	Patients reject AI tool	Delays in software rollout
	3	Potential Cause	Unfriendly user interface	Concerns over safety and privacy	Training, software, or hardware issues
	4	Severity	Minor	Minor	Minor
	5	Probability	Occasional	Frequent	Frequent
	6	Hazard Score	3	4	3
	7	Action	Control	Accept	Control
	8	Description of Action	1. Educate staff on tool benefits and provide significant opportunities for discourse 2. Develop and implement a significant training regimen 3. Use incentives to motivate staff to use tools	1. Educate the patient on the tool's functionality and answer questions 2. Respect patient's wishes not to receive AI-powered healthcare 3. Note the patient's wishes in their chart to prevent future difficulties	1. Coordinate with IT and staff to develop a software implementation plan 2. Design timelines with room for error 3. Ensure open channels of communication between various departments for setbacks or other errors

Table A4

FMEA Process Step 4 – Monitor Quality and Reliability

Process Step #4	1	Process Step	Monitor Quality and Reliability		
	2	Potential Failure Mode	Staff do not provide accurate feedback	Staff do not use AI tools per training	Staff reject AI tools
	3	Potential Cause	Fear and/or dislike of AI-based technology	Confusing user-interfaces	Inaccurate or unclear model outputs
	4	Severity	Moderate	Major	Minor
	5	Probability	Frequent	Frequent	Frequent
	6	Hazard Score	8	12	4
	7	Action	Control	Control	Control
	8	Description of Action	1. Incentivize honest feedback 2. Use quantitative metrics to gauge tool performance 3. Provide statistical and software education to staff to improve technical literacy	1. Provide regular training 2. Work with developers to improve software usability 3. Garner feedback from staff	1. Provide regular training 2. Communicate staff feedback with developers 3. Work with developers to improve model transparency and facilitate model augmentation

Table A5

FMEA Process Step 5 – Report Errors

Process Step #5	1	Process Step	Report Errors		
	2	Potential Failure Mode	Staff cannot access reporting tools on QMS platform	Reports are not databased properly	Reports are not flagged for review properly
	3	Potential Cause	QMS platform servers are under maintenance	Errors in report submission, Invalid responses	Inadequate description of the issue
	4	Severity	Minor	Minor	Major
	5	Probability	Occasional	Frequent	Occasional
	6	Hazard Score	3	4	9
	7	Action	Control/Eliminate	Eliminate	Eliminate
	8	Description of Action	1. Control: Schedule maintenance during non-peak hours. 2. Ensure users are alerted to planned maintenance well in advance. 3. Eliminate: Ensure significant redundancy in QMS-hosting servers	1. Ensure software can handle variations in data entry. 2. Communicate errors in data entry to the user before sign-off. 3. Save data as it is entered. 4. Choose no-SQL databases (if possible).	1. Provide training on reporting best practices. 2. Champion a culture of quality. 3. Ensure that reports are flagged appropriately by reviewers